Collection of
Technical Papers
on

Guidance and Control

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

TO.
A.C. LATTANZI
FROM.
K. TSUTSUMI

Collection of
Technical Papers
on

Guidance and Control

PRICE:
MEMBER $16.00
NONMEMBER $32.00

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

American Institute of Aeronautics and Astronautics / 1290 Sixth Avenue, New York, N. Y. 10019 / 212 LT 1-4300

AIAA GUIDANCE, CONTROL, AND FLIGHT DYNAMICS CONFERENCE COMMITTEE

General Chairman
JESS B. HUFF
U. S. Army Missile Command

Technical Program Chairman
RICHARD P. GILLOOLY
McDonnell Douglas Corp.

Administrative Chairman
FRED E. DIGESU
NASA/Marshall Space Flight Center

PROGRAM COMMITTEE

MALCOLM J. ABZUG
TRW Systems

BARRY W. BOEHM
The RAND Corp.

FRANZ T. GEYLING
Bell Telephone Labs

WILLIAM A. GRIFFITH
U. S. Army Missile Command

WALTER L. PONDROM JR.
North American Aviation Inc.

Assistants to Administrative Chairman
JOHN CAMPBELL
Pratt & Whitney Aircraft

DOOLEY CULBERTSON
Thiokol Chemical Corp.

Chairman of Coordinators
JOHN VISLOCKY
IBM Corp.

Public Relations
ARTHUR BENNETT
Boeing Co.

Transportation
GEORGE DETKO
NASA/Marshall Space Flight Center

WILL LEWIS
U. S. Army Missile Command

Housing/Facility
WILLIAM EARNSHAW
Raytheon Company

Membership
CHARLES THOMSON
ITT

Security
WILLIAM FERGUSON
NASA/Marshall Space Flight Center

Coordinators-of-the-Day

ARTHUR BENNETT
Boeing Co.

H. B. VALLEY
IBM Corporation

P. H. BUTLER
Chrysler Corp.

INDEX

INDEX

No. 67-532

ATTITUDE DETERMINATION AND CONTROL OF THE SYNCOM, EARLY BIRD, AND APPLICATIONS TECHNOLOGY SATELLITES

by

W. H. SIERER and W. A. SNYDER
Hughes Aircraft Company
El Segundo, California

AIAA Paper
No. 67-532

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

7.01, 7.03

-- NOTES --

ATTITUDE DETERMINATION AND CONTROL OF THE SYNCOM, EARLY BIRD, AND APPLICATIONS TECHNOLOGY SATELLITES

By

W. H. Sierer and W. A. Snyder
Hughes Aircraft Company
Space Systems Division
El Segundo, California

Abstract

The Syncom, Early Bird, and Applications Technology Satellite are spin-stabilized, synchronous orbit communications satellites with hydrogen peroxide attitude and velocity control systems and an apogee kick motor for insertion into final orbit. Precise attitude determination and control are required for alignment of the apogee kick motor prior to apogee boost and in the final orbit where the spin axis is maintained normal to the orbit plane for communications.

This paper discusses attitude determination, including the data types and the estimation process, and the use of the hydrogen peroxide control system in changing spacecraft orientation. Next a normal launch sequence is covered, and the interrelation between attitude determination and control during critical mission phases, including real time analysis during reorientation, is explained. Determination and control for final orbit operations, including the long term effect of solar radiation pressure precession, is then discussed. Finally the accuracies achieved by current methods as indicated by operational experience and new techniques in attitude determination are covered.

I. Attitude Determination

Attitude is determined on the ground by telemetered output from on-board sun sensors and by measuring the plane of polarization of the down-link signal from the spacecraft communications antenna. Each data type can be used independently for determination of the spin axis direction, and each takes advantage of unique spacecraft characteristics to allow simple and reliable, yet very accurate, attitude determination.

Each sun sensor consists basically of a photovoltaic element behind a narrow slit aperture. Thus the device, which is mounted on the periphery of the spacecraft, has a fan-shaped field of view which is scanned through 360° by the satellite spin. A measure of the angle between the spin axis and the spacecraft-to-sun line is obtained by mounting two such sensors together, one oriented such that the plane of its fan-shaped field of view contains the spin axis, and the other mounted such that its beam makes an angle of 35° with that of the first. (See Figure 1.) Inspection of this geometry indicates that in general the two sensors will not "see" the sun simultaneously but will be triggered some time interval apart (except when the sun lies along their line of intersection which is 90° to the spin axis). This time interval may be converted to an equivalent angle through which the spacecraft rotated between pulses since the spin period is known. This angle, designated ψ_2, is related to ϕ, the angle between the spin axis and the spacecraft-to-sun line, by the equation

$$\cot \phi = \cot 35^\circ \sin \psi_2 .$$

Thus a single ψ_2 measurement defines a cone of half angle ϕ centered on the spacecraft-to-sun line on which the spin axis must lie.

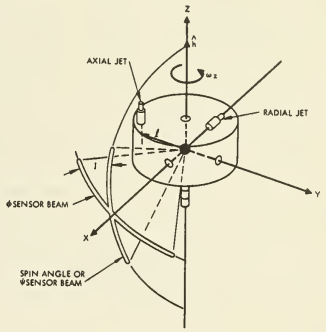


FIGURE 1.
SUN SENSOR AND CONTROL JET GEOMETRY

The second source of data relating to spacecraft attitude is the linearly polarized radiation from the spacecraft communications antenna. Mounted at the opposite end of the satellite from the apogee motor, this antenna is colinear with the spin axis, so that the electric field vector of the radiated signal lies in the same plane as the spin axis. At a ground terminal, measurement of the angle between the plane defined by the local vertical and the line of sight to the satellite and the plane of polarization of the down-link signal yields data which, when coupled with station location and orbital parameters, provides an independent indication of spin axis attitude.

The plane of polarization is subject to Faraday rotation by the ionosphere, thus complicating its use. However, this effect can be analytically described with knowledge of the signal's frequency, the earth's magnetic field, and the electron density of the upper atmosphere. Of these the last is least certain.

If both a ψ_2 reading and a polarization angle measurement are taken simultaneously, the cone and plane previously described will intersect in general along two lines, one of which is the spin axis direction. Additional measurements (of either type), separated in time from the initial pair, will

eliminate the ambiguity. In fact, two data readings of the same type if separated in time will define the spin axis direction. It should be noted, however, that the time separation needs to be substantially longer for sun data since the ϕ angle changes as a result of the earth's orbital motion about the sun, while the more rapid orbital motion of the spacecraft relative to a ground station has a much more immediate effect on polarization readings.

In practice, numerous data of both types are collected over whatever time span is available, this ranging from several hours during the launch phase to weeks or months after the satellite is on station. Isolated pairs of data or short period averages of data are normally used only for "quick look" attitude evaluation during spin axis reorientation maneuvers.

The attitude determination system includes equipment at the ground terminal for measuring polarization angle and for decoding sun sensor pulses from spacecraft telemetry, data transmission to the operations and control center, and a large scale digital computer with an attitude determination program capable of processing the data and estimating attitude. The computer program currently used features a weighted least squares estimation procedure which differentially corrects an a priori estimate of attitude and its covariance matrix using the most recent data. A detailed ephemeris of the satellite, including the perturbing effects of the earth's oblateness and triaxiality and of the sun are used, and a dynamic model of solar radiation pressure precession of the spin axis (discussed in a later section) is included. Faraday rotation effects are modelled and during launch operations the inclusion of the latest day-to-day measurements of the ionospheric plasma frequency (or critical frequency) are used to determine electron density.

As implied by the weighted least squares technique used, data weighting is important and should reflect the confidence in the data types. Sun sensor data are more accurate, with experience indicating a root-mean-square deviation of slightly less than 0.1° in most cases. The quality of polarization data varies from station to station, with rms deviations usually less than 1°. However, a phenomenon peculiar to this data type is measurement bias due to equipment inaccuracies. The magnitude of such bias varies from station to station and is difficult to detect prior to launch. To alleviate this problem, station polarization angle measurement biases may be included as unknowns to be determined in the differential correction procedure. This technique has been successfully employed and has enhanced data reliability.

Biases in sun sensor (ψ_2) data originate predominately on the spacecraft and reflect misalignment between the sensors and the spin axis. These are negligible and are to be ignored.

II. Attitude Control

A spin stabilized spacecraft whose roll-to-pitch inertia ratio is greater than unity will maintain its spin axis fixed with respect to inertial space in the absence of external torques in accordance with the law of conservation of angular momentum. Since a detailed discussion of the dynamics of a spin-stabilized configuration is contained in References 1, 2, and 3 only a brief description with emphasis on the control aspects* is presented here for completeness.

The hydrogen peroxide control system consists of two sets of body-fixed axial and radial jets. The radial jets fire outward (perpendicular to the spin axis) through the center of gravity and when pulsed in synchronism with the spin speed allow vehicle velocity corrections to be performed. (See Figure 2b) The axial jets, aligned parallel to the spin axis, are mounted near the periphery of the spacecraft on the apogee motor end.

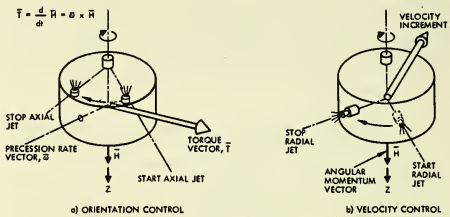


FIGURE 2.
PULSED JET CONTROL FOR SPIN-STABILIZED SPACECRAFT

In performing a reorientation maneuver a precession torque is applied to the spinning spacecraft by one of the body-fixed axial jets which is pulsed in synchronism with the spin speed over a given sector of the spin cycle (Figure 2a). In accordance with Newton's second law of (angular) motion,

$$\overline{T} = \frac{d}{dt}\overline{H} = \overline{\omega} \times \overline{H}, \quad \overline{H} = \text{constant}$$

where

$\overline{T}$ = applied torque vector,

$\overline{H}$ = angular momentum vector, and

$\overline{\omega}$ = precession rate vector.

In this pulsed jet method of generating torque the vectors $\overline{T}$, $\overline{H}$, and $\overline{\omega}$ are mutually perpendicular.

It can be shown (Reference 4) that the nutation induced by the axial jet during the pulsing mode is bounded by a small angle ($\sim 0.1^{\circ}$) for a reasonable spin speed ($\sim$100 RPM) and roll-to-pitch moment of inertia ($\sim$1.2). Thus, for all practical purposes the motion of the geometric spin axis follows that of the angular momentum vector during the precession process.

The reference signal for synchronous pulsing is obtained from the sun sensor. Knowledge of the ephemeris of the sun and the spacecraft orbital elements is sufficient to generate spin phase information for the axial and radial jets relative to the sunline. Thus with very simple spacecraft sensor instrumentation and body-fixed thrustors, both attitude and velocity control are provided. Alignment of the jets relative to the spacecraft center of gravity and spin axis is not too critical

* All based on the original work of D. D. Williams, Reference 1.

for thrust levels of the order of one to five pounds because of the gyroscopic "stiffness" afforded by the spinning spacecraft, the primary undesired effect of misalignment being a change of spin period.

III. Launch Operations

This section has a three-fold purpose. First, it illustrates and serves to tie together the theory previously discussed. Secondly, it brings out the hard, cold realities that force adjustments, refinements and compromises in applying techniques. Finally, certain real time techniques used in spacecraft control are introduced.

For purposes of illustration, we choose to follow the launch of the Intelsat II-C for the Communications Satellite Corporation on 23 March 1967. The spacecraft was launched from the Eastern Test Range with a Thrust Augmented Delta booster at 0130 Universal Time. The solid propellant third stage injected the satellite into a near perfect elliptical transfer orbit with apogee over the equator and near the synchronous orbit altitude of 19,310 n. mi., with perigee height near 180n.mi., and with orbital inclination of 26.3°.

The initial spin axis attitude shown on the attitude history chart (Figure 3) was near that of the spin stabilized third stage after its firing, and, as the vehicle traversed the transfer orbit, this orientation was maintained in inertial space. The satellite was observable for tracking on successive even numbered apogees for periods of several hours during which a sequence of reorientations were performed to bring the spin axis as close as possible to the desired attitude for fourth stage firing.

The initial attitude was soon determined to be near that expected from nominal third stage performance. During the first hours of ground station visibility several hundred data values were received. Because the relative motion of the sun during this short time interval was so slight, there was little change in the solar aspect angle, Ø, and all the sun sensor data contained essentially the same information. Thus, polarization data, though subject to the inaccuracies previously described, was essential to uniquely define the spin axis direction within the short time intervals available.

Having determined the initial spacecraft attitude, reorientation of the spin axis to an attitude offset by 5° from the desired attitude for apogee motor firing was then planned. Maneuver commands and pulse phasing information appropriate to the time of reorientation and the ground station executing the maneuver were generated, and a chart similar to that shown in Figure 4 was prepared. On this chart spin axis attitude is mapped in a coordinate system directly related to the received attitude data as contrasted to the inertial right ascension and declination coordinates used in Figure 3. Thus, a reorientation maneuver lasting several minutes can be monitored by using intermediate values of the data to plot instantaneous attitude.

Because the phasing of the axial jet pulses during a reorientation maneuver is constant relative to the sun, the spin axis describes a loxodrome or rhumb line in spherical coordinates having the sun as a pole; on the chart used, which is a Mercator projection of that sphere, the loxodromic motion can be plotted as a straight line.

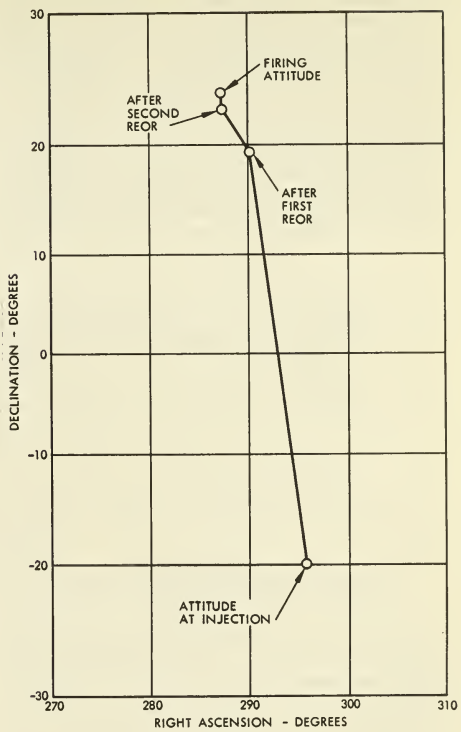


FIGURE 3
SPIN AXIS ATTITUDE

Contours of constant polarization angle are approximated on the chart by straight lines, as shown in Figure 4. While the locus of spin axis attitude on the chart changes quite slowly with time (as the sun moves), the polarization contours are valid only for a particular ground terminal at a particular time. Thus in Figure 4, two sets of polarization contours are shown, one for 1900Z, and another for 1915Z. The first leg of the reorientation maneuver was executed at 1900Z and left the spin axis at the location shown, about midway to the target attitude. The remainder of the maneuver, with corrected phase and magnitude, was executed at 1915Z and left the spin axis about 4° short of the objective. It may be noted that the second set of polarization contours, by which the final spin axis attitude was deduced, are translated somewhat with respect to the first set.

Errors in reorientation are attributable to errors in pulsing phase and magnitude. Phase errors result from imperfect knowledge of electronic delays associated with commanding the spacecraft and uncertainty in the pulse centroid of the peroxide system, while magnitude errors result from imperfect calibration of system performance stemming

from hydrogen peroxide's inherent chemical instability, and varying storage tank pressures and temperatures. Both phase and magnitude errors can be determined graphically very quickly in real time using the chart. If the error is large enough to warrant a further maneuver, or touch-up, to correct it, the parameters for the command instructions can be obtained directly from the chart. Touch-up maneuvers are generally restricted to follow large reorientations (>10°) where a phase error of several degrees or a magnitude error of 3% - 6% would leave an error larger than the uncertainty of the data with which that error was determined. It should be noted that "in-flight" calibration of system performance can considerably reduce pre-flight calibration uncertainty.

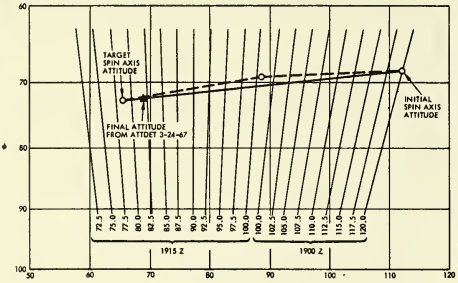


FIGURE 4
REORIENTATION PLOTTING CHART

The initial large reorientation appeared to indicate system underperformance with a phase error of about 3°. However, since the target had been deliberately offset from the apogee motor firing direction to obtain a more favorable sun aspect from the standpoint of spacecraft temperature, no immediate touch-up was planned. Subsequent collection and computer processing of additional data confirmed the new spin axis direction, and a further reorientation during the next visibility period was planned and executed. A third very small correction was required just prior to apogee motor fire to accommodate a revised firing target attitude.

Mention should be made of an additional method of confirming attitude. If a velocity correction is made to the orbit, it is possible to infer the velocity vector actually required to get from the initial to the final orbit. Knowledge of this velocity vector indicates spacecraft attitude. While limited by orbit determination uncertainty, this technique is useful for velocity corrections greater than about 100 feet per second. The apogee boost, which delivers about 6000 feet per second, thus generally permits a very accurate evaluation of the actual attitude during apogee motor firing. While not done in the Intelsat-C launch, a velocity correction of up to several hundred feet per second prior to boost is sometimes made using the hydrogen peroxide system, thus allowing an assessment of attitude prior to motor firing when it is more critical. In the Intelsat-C launch, the attitude indicated by the apogee motor firing was less than 0.3° from that determined by the attitude data.

IV. Operational Considerations and Solar Precession

After apogee motor firing, the spin axis is reoriented normal to the orbit plane so as to direct the radiation pattern of the communications antenna toward the earth. In the course of normal operation the spacecraft is not maneuvered except to counter the effects of orbital perturbations. (See Reference 5). Depending on satellite location and station keeping requirements, maneuvers are performed at intervals of several weeks to several months. These velocity corrections can induce small attitude changes due to jet misalignment, and, depending on pointing requirements and the frequency of these corrections, small attitude reorientations may be required.

A more insidious form of attitude perturbation is precession due to solar radiation pressure. The effect of solar radiation pressure is well-known to those concerned with precise orbit determination. However, the drum-shaped satellite in general will not have its center of gravity coincident with its center of pressure. Thus in addition to introducing an orbital perturbation, this effect will torque the satellite, precessing it in a systematic fashion.

Examination of the dynamic situation reveals a source of torque independent of the center of gravity offset. For a near-equatorial orbit, the spin axis is approximately aligned with the earth's axis. The seasonal motion of the sun then will slowly change Ø, the angle the sun makes with the spin axis, up to a maximum of ±23° from the 90° position during the course of the year. Toward the extreme values of Ø, shear torques on the spacecraft sides become important. (See Figure 5).

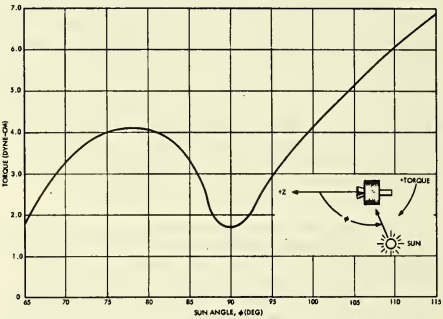


FIGURE 5
TORQUE vs SUN ANGLE
HS-303A

By modeling the reflective properties of the spacecraft, the values of torque shown in Figure 5 as a function Ø were computed. Numerically integrating the vector equations of motion using this torque curve and starting with the initial spacecraft attitude allows prediction of future (or past) attitudes and, when used in the attitude determination program as previously described, allows a more accurate fit of observed data. An example of attitude prediction is shown in Figure 6. Note that the instantaneous direction of precession must always be perpendicular to the satellite-to-sun line.

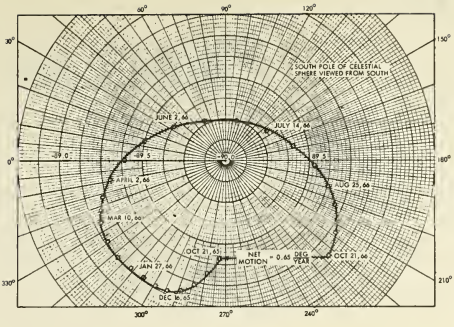


FIGURE 6
HS-303A SPIN AXIS PRECESSION
Angular Momentum=168.3slug $\frac{ft^2}{sec}$

In general the predicted motion will result in a net precession over the course of a year of up to several degrees. The exact shape and magnitude of the precession curve depend on individual spacecraft characteristics. Normally this precession would have to be corrected with a reorientation maneuver to maintain the spin axis normal to the orbit plane, but a surprising assist from the sun and the moon reduces this requirement. The orbit normal itself is precessed approximately 0.9° per year by solar and lunar perturbations in the same direction as the spin axis motion. Thus to the extent of a similar magnitude of spin axis precession, the spin axis can be passively controlled by solar radiation pressure. This could even be a design criterion on future spacecraft.

V. Determination Accuracy and Future Systems

As previously discussed, both sun angle data and polarization angle data are required for complete specification of attitude during the launch phase. Accuracy, then, is limited by the less accurate polarization data with its rms deviation of around one degree. In theory, for a gaussian or purely random distribution of measurement errors the uncertainty in attitude should decrease as $N^{-\frac{1}{2}}$ where N is the number of measurements. This assumption is implicit in the weighted least squares method of estimation. However, the possibility of bias in these measurements dominates the certainty of the solution. Even in the solution mode where station biases are estimated as unknowns, the geometry is such that a unique solution is not always possible. Thus a working uncertainty for apogee motor firing on the order of 0.5° to 1° is used to ensure conservatism in that crucial analysis. Recall the 0.3° error for Intelsat II-C as previously discussed.

In final orbit operations, sufficient time is available to allow determination with sun sensors only. In this case, accuracy to within several hundredths of a degree is indicated based on individual data rms deviations of 0.1°. However a discouraging word must be entered here. Alignment tolerances for the sensor itself are such that 0.1° to 0.2° error from this source is possible. In theory, it is possible to estimate such alignment errors in the least squares process but analysis shows that six months to a year's data from an unperturbed satellite would be required; this is not likely to become available.

There are several other methods for obtaining attitude data from a spin stabilized satellite. One which has been used with some success is comparison of received signal strength from the spacecraft communications antenna with the known radiation pattern. Since the antenna and its radiation pattern have a known spatial relationship to the satellite, attitude can be inferred.

A method which in theory would be very a accurate is a star scanner. Such a device, probably a photomultiplier tube, would be mounted roughly parallel to the spin axis but on a moveable track. Pointing "north", Polaris or another suitable star could be found as a blip each revolution and the offset angle required to "see" it would indicate attitude. Such a system, when compared to current techniques, is considerably more expensive and certainly less reliable but could offer increased accuracy.

The most promising new technique is that of the infrared horizon crossing indicator or earth sensor. Such a device scans by virtue of the satellite's spin and sends a signal each time it goes on or off the infrared horizon of the earth. Knowing the offset angle between the spin axis and sensor optical axis and the proportion of the spin cycle during which the earth was in view would determine the angle between the local vertical and the spin axis. Error sources for this device are alignment errors, electronic signal processing errors, and uncertainty in the earth's infrared horizon. Recent analysis shows that the sum of such errors can be kept below 0.1°; further, since the earth sensor sees its full cycle of measurements in one day (as opposed to one year for the sun sensor), least squares estimation of certain alignment and signal processing biases is possible. This technique could result in attitude determination accuracies to several hundredths of a degree with off-the-shelf equipment.

VI. References

1. Williams, D.D., "Dynamic Analysis and Design of the Synchronous Communication Satellite", Hughes Aircraft Company TM 649, May 1960.
2. Syncom II Summary Report, Hughes Aircraft Company SSD 3128R (Revised) NASA Contract 5-2797, 31 March 1963.
3. Summary Report, Advanced Syncom, Vol. 1, Hughes Aircraft Company SSD 31118R, NASA Contract 5-2797, 31 October 1963.
4. Neufeld, M. J., "Launch and Orbital Analysis", Hughes Aircraft Company, SSD 3107R, NASA Contract 5-2797, October 1963, pp 3-52.
5. Neufeld, M. J. and B. M. Anzel, "Synchronous Satellite Stationkeeping", Communications Satellite Systems Technology, Ed. by R. B. Marsten, Academic Press, 1966.

67-532

-- NOTES --

No. 67-533

LUNAR ORBITER ATTITUDE CONTROL SYSTEM

by

DOUGLAS C. FOSTH and WARREN I. MITCHELL
The Boeing Company
Seattle, Washington

AIAA Paper
No. 67-533

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

2.03, 7.01, 7.03

-- NOTES --

LUNAR ORBITER ATTITUDE CONTROL SYSTEM*

Dr. Douglas C. Fosth
Research Specialist

Mr. Warren I. Mitchell
Senior Group Engineer

Space Division
THE BOEING COMPANY
Seattle, Washington

Abstract

This paper on the Lunar Orbiter Attitude Control System is divided into two major parts covering the design of the system and the space flight performance of the first three flight systems. The section on the system design discusses the Lunar Orbiter mission and gives the guidance and control requirements derived from the mission. System design criteria, concept trades, and a functional description of the final system choice are given. The functional description includes a complete block diagram as well as block diagrams and phase plane plots covering the six major system functions; Sun acquisition and tracking, Canopus acquisition and tracking, attitude maneuvering, powered flight control (TVC), ΔV maneuvering, and inertial hold. The second part presents typical data covering the major system functions described in the first part, and compares the space flight data with ground test data. Included are both system and component data -- particularly Inertial Reference Unit and Canopus Star Tracker component data. Two space flight attitude control problem areas, Canopus acquisition and thermal control, are discussed along with the hardware and software modifications necessary for solution of these problems.

Introduction

This presentation covers the design, implementation, and flight of the Lunar Orbiter Attitude Control System. The Lunar Orbiter mission is described by mission phase and the associated attitude control requirements are given. The design concept of the Attitude Control System (ACS) is given, followed by a functional description of the ACS implementation and mission performance.

Mission Description

The trajectory of the spacecraft to accomplish the mission objectives is shown in Figure 1. The vehicle is launched from Cape Kennedy by an Atlas-Agena D booster combination and placed in orbit about the Earth. At the proper place in the orbit the Agena is restarted and boosts the spacecraft to a translunar trajectory. During this portion of the mission the ACS is essentially inactive with the nitrogen kept from the thrusters by normally closed squib valves. As a protective measure for the Inertial Reference Unit (IRU), the gyros are operating in the rate mode. This is the safest condition for the gyros during a boost environment. The Sun Sensor is disconnected from the electronics to minimize the number of commands to turn the thrusters on when no gas is available for cooling and lubrication. The Star Tracker is turned off to prevent damage from high voltage arcing during the transition from Earth atmosphere to the hard vacuum of space and during the passage through the Van Allen belts.

Following separation from the Agena booster, the spacecraft deploys the mechanical equipment (solar panels and antennas), fires the nitrogen squib valves, connects the Sun Sensor and acquires the Sun. A worst case rate from separation was assumed to be three (3) degrees per second and acquisition is required within 20 minutes of separation in order to not deplete the batteries. After approximately six (6) hours, the spacecraft goes through a programmed acquisition of the star Canopus for the Roll Axis attitude reference. During the nominal translunar flight there is one planned midcourse correction to realize the proper rendezvous with the Moon. (More are possible if required.) The attitude maneuver requirements and the velocity change measurement are not particularly critical during this correction and do not place any stringent requirements on the system.

At the Moon, the spacecraft is placed into an elliptical orbit with a nominal 250 KM perilune and 1850 KM apolune. The injection sequence provides operational requirements for the attitude maneuver accuracy and ΔV measurement. The engine burn during this phase is the longest of the mission and poor performance here would be most detrimental to the mission performance. Orbital and gravitational data as well as high altitude- low resolution pictures are the primary objectives of this portion of the mission. When sufficient data have been gathered about the orbital characteristics of the spacecraft, the orbit is changed to a different ellipse with a perilune of 45 KM and the same apolune of 1850 KM. In the performance of this transfer maneuver, angular accuracy is not critical, but the ΔV measurement requirement is stringent. When the time arrives for taking pictures, an attitude maneuver is made during each orbit at sunrise in order to orient the spacecraft (and camera) for the photo-taking sequence.

System Requirements

Figure 1 shows detailed functional requirements placed upon the system, correlated with the mission phases. These requirements have been placed upon the ACS as guidelines for hardware procurement, system design, and to allow studies to be made of mission accomplishment requirements. The specific requirements which controlled hardware specifications are given in Table 1, both the original requirements and those generated later in the program as necessary to accomplish the overall mission objectives. The guiding design philosophy for Lunar Orbiter has been to use proven system concepts and to implement them with available qualified components to increase reliability and to save time and money. Where necessary components were not available in qualified form, care was taken to assure the timeliness of necessary development and to minimize the need for extending the state of the art beyond the immediate and necessary requirement.

*Accomplished by Boeing under NASA Contract NAS-1-3800, LANGLEY RESEARCH CENTER, Hampton Virginia.

67-533

		INITIAL	FINAL
Maneuver angle error (RSS) (up to 180° maneuver)	%	±0.32	±0.32
IRU rate error (@ 0.55 ± 0.05°.sec)	%	±0.10	±0.11
Integration error	%	±0.30	±0.30
Velocity change error (% of total ΔV)	%	±0.50	±0.90
Accelerometer error (@ 0.09 to 0.22 g)	%	±0.07	±0.40
Integrator error (% of total ΔV)	%	±0.23	±0.23
Thrust tailoff error (% of total ΔV)	%	±0.03	±0.06

Table 1 OPERATIONAL REQUIREMENTS

The design criteria for the ACS are shown in Table 2.

ESTABLISH AND HOLD SUN AND CANOPUS ATTITUDE REFERENCES

MANEUVER ACCURATELY FROM CELESTIAL REFERENCES

HOLD ATTITUDES INERTIALLY FOR
- o ΔV MANEUVERS
- o LUNAR PHOTO SEQUENCES

POINT HIGH-GAIN ANTENNA

PROVIDE ΔV MEASUREMENT & CONTROL (TVC) DURING BURN

PROVIDE SOLAR ORIENTATION FOR UP TO ONE YEAR

Table 2 DESIGN CRITERIA

Design Concept

The overall system design is summarized in Table 3.

1. On-board digital programmer, within the Flight Electronics Control Assembly, directs all spacecraft activities.
2. Orientation in space referenced to inertially fixed stars.
3. Three axis stabilized configuration, which orients solar panels to Sun, maintains a stable base for pointing of camera, antenna, and direction during engine firing.
4. Maneuvers measured by an inertial reference unit (IRU), i.e., strapdown rate-integrating gyros.
5. Velocity control provided by integrating the output of a linear accelerometer.
6. Vehicle angular torques provided by cold gas reaction jets during normal operation.
7. Vehicle angular torques provided in Pitch and Yaw by thrust vectoring, and in Roll by cold gas reaction jets, during velocity engine burn.

Table 3 DESIGN CONCEPT

The use of an onboard digital programmer to direct the spacecraft activities is necessitated by the fact that the spacecraft must perform many functions when it will not be possible for an Earth tracking station to communicate with the craft. The memory of the Programmer is used to store the programs which are to be performed and also to serve as a part of the Programmer's registers while it is performing its operating functions. The stored program capability of the Programmer is at least sixteen hours and this time is largely dependent upon the tasks to be performed. The method chosen to load the Programmer is to read a program into the memory and then to verify it through a telemetry readout of the Programmer memory.

The Closed Loop Electronics, used for the analog operation of the ACS, is packaged as an integral part of the Control Assembly, with a common power supply and - in several instances - sharing the use of common inputs.

The orientation of the spacecraft in space is with the Roll axis pointed to the Sun for Pitch and Yaw control and the Pitch axis to the star Canopus for the Roll control. The orientation to the Sun is used primarily to provide radiation power to the solar panels. Secondary reasons are the availability of reliable sensors for that celestial reference, and the simplification of the thermal control problem with that particular orientation. (Canopus was used as a reference for Mariner spacecraft and development work had progressed on advanced sensors for that star to the point where a sensor with performance meeting Lunar Orbiter requirements would be ready for the Lunar Orbiter program.)

The Earth was not considered as a reference because of launch window constraints and the large occultation periods which could occur each orbit.

The ΔV which is imparted to the spacecraft is controlled by digital integration of the measured acceleration of the vehicle. The acceleration is measured by a pulse rebalance accelerometer which lends itself very nicely to digital integration. This method of controlling the resultant ΔV on the spacecraft allows a much tighter control to be exerted over the midcourse maneuver and the two orbital injection maneuvers.

The vehicle angular torque is provided to the spacecraft by a cold gas reaction jet system during normal system operation. The cold gas system was chosen because of availability of components with known reliability, proven performance and low weight for the mission desired.

During the main engine firing, the control torques about the Pitch and Yaw axes are provided by engine gimballing, using an electric motor thrust vector control system. The thrust vector control gimbals and servos are not discussed in any detail in this paper, even though the command signal originates in the IRU and is processed in the Control Assembly. It will be sufficient to say that this method of control has the highest reliability, lowest weight, and shortest development time.

System Description

The general block diagram of the ACS is shown in Figure 2. Hardware locations on the vehicle are given in Figure 3. A detailed description of the system which considers all phases of mission operation is given below.

The system is basically a rate limited conventional ON-OFF gas reaction control system. The rate damping signal is provided by a rate gyro for

each axis during normal operation with the Sun Sensor (Pitch and Yaw) or Star Tracker (Roll) not occulted. For the primary photo mission, the deadband can be commanded to either 0.2 or 2.0 degrees, depending on the particular operating phase in use. For the extended mission, the Star Tracker is turned OFF and the coarse Sun Sensors (wide angle) are switched in to increase the two deadbands on those axes to either 1.2 or 12.0 degrees. The rate to position gain that has been chosen is 4:1 which gives end point motion for most operating conditions.

With the Sun or Canopus occulted, a lead-lag compensation network is used in an inertial hold mode for each axis. The network obtains position error signals from the gyro in rate-integrate mode. The same rate-to-position gain is used with a filter cutoff of ten radians to minimize the noise effect into the system.

The ON command to the thrusters is through an eleven millisecond "one-shot" circuit in parallel with a direct wire to the thruster driver to provide no ON times of less than eleven milliseconds. This minimizes the effect of noise on the operating life of the thruster and improves the limit cycle performance.

An additional optional input is provided for the Yaw axis during photographic periods in the form of a V/H or Crab Angle input. This signal is used to reduce the smear on the photographs by decreasing the error between the Camera IMC axis (parallel to S/C Roll axis) and the orbital velocity vector which minimizes the cross axis velocity component.

The ACS has five basic modes of operation which are used in different combinations during the flight. They are:

Celestial Reference Acquisition
Limit Cycle
Maneuver
Inertial Hold
Rate Limit

The use of these modes will be shown as the different phases of the flight are considered.

Sun Acquisition

The primary Sun Acquisition is accomplished according to Figure 4. For this initial acquisition, the Sun Sensor field of view is 4π steradian. The output from the Sun Sensor amplifier is limited with a soft limiter. This attitude error limiter has the effect of a rate limit on the system. Depending upon the deadband chosen, the rate range is nominally 0.05 to 1.05 degrees per second for the 2.0 degree deadband and 0.50 to 0.60 degree per second for the 0.2 degree deadband. For the Sun acquisition, the time allowed from Agena separation is limited to twenty minutes by battery capacity to power the vehicle for that period. Hence the narrow deadband is chosen for this task to assure that the acquisition rate will be at least 0.5 deg. per second.

In spite of the need for fast acquisition, rate limit is used for fuel saving. Even with the 'optimum' type of bang-bang system and perfect switching, the fuel consumed from a large initial rate and/or position error would be much larger than with the rate limited concept chosen. Using the rate limit scheme, the time to acquisition is within the allowed twenty minutes and the fuel consumption is moderate and dependent upon the separation rate condition.

During this acquisition, which is prior to Canopus acquisition, the Roll axis is operating in a rate limited mode (IRU Roll channel is in Rate Mode), the position signal is zero (Canopus Star Tracker is turned OFF), and therefore the Roll rate deadband is set at ±0.05 degree per second by the use of the 0.2 degree deadzone in the Roll channel.

After four to five hours, the Pitch and Yaw deadbands are opened to the large deadband (2.0°), the fine sensors are turned ON and the coarse sensors are turned OFF. With the Sun acquired, this places the spacecraft in the ±2.0° limit cycle mode about the Pitch and Yaw axes.

Canopus Acquisition

Approximately six hours after launch, the Canopus Tracker is turned ON and the Canopus Acquisition sequence is initiated as shown in Figure 5. A precise 360° Roll maneuver is commanded during which time a star map is telemetered to Earth. This star map is compared with an overlay of the predicted star map signal (see Figure 6). This overlay will then accurately define the Roll angle of the craft relative to Canopus at the end of the 360 degree maneuver. A second maneuver is commanded to orient the S/C to Canopus with a small negative Roll error relative to Canopus. The third maneuver in the sequence is an Acquire Canopus Plus command which rolls the spacecraft in a positive direction at 0.55 ±0.05°/sec. until Canopus Recognition is received by the Control Assembly from the Tracker. Upon receipt of the recognition signal, the maneuver slew is terminated and the spacecraft acquires the star (the Tracker acquires Canopus and generates the recognition signal as soon as Canopus enters the ±4.1° Roll field of view.) The Roll axis will then be in the limit cycle mode with attitude error from the Tracker and rate error from the IRU. It continues this way until the maneuver sequence for the midcourse trajectory correction is initiated.

Maneuver

The Lunar Orbiter maneuver sequence is different from any method used to date in a spacecraft ACS. The sequence is outlined in detail in Figure 7. For the maneuver sequence, the deadbands for all axes are first narrowed to 0.2 degree. This provides tighter control of the vehicle rate during the maneuver and also guarantees the vehicle position to be within 0.2 degree of the celestial reference at the start of the sequence.

After a wait time to stabilize the new limit cycle, the maneuver is initiated by simultaneously applying a 0.55 ±0.05 degree per second slew voltage to the maneuver axis, and switching the IRU rate signal to a voltage-to-frequency (V/F) pulse converter which in turn is connected to the pulse counting register of the Programmer. The control system will null the slew voltage by firing the proper jets to bring the S/C to commanded rate. The integration register in the Programmer will count the output pulses of the V/F converter and compare the accumulated count with that stored in the memory for the commanded maneuver.

When the maneuver has been completed, the slewing signal is removed and the gyro is switched to the rate integrating mode. This fixes the zero of the inertial hold mode to the desired end point of the system maneuver. The vehicle is controlled during the maneuver hold period with jet pulses commanded by the inertial hold compensation network operating on the position output from the gyro.

Design Studies

The selection of the techniques that were chosen to implement these operating schemes was based on several factors. The literature and technical reports of the experience of the many existing satellites were studied. The shortcomings and the benefits that existed for each of the techniques were weighed to determine their effect on the Lunar Orbiter mission. The results of laboratory studies made with company funds and company funded research efforts were drawn on for help in making the decisions.

The different methods which could be used for compensation during the inertial hold phase were given extensive study. The methods which were considered are the following:

High gain gyro with capacitor feedback for position, and lag around the switching amplifier for pseudo-rate signal.

High gain gyro with resistor-capacitor feedback for a position plus rate signal.

Low gain gyro with lead-lag compensation for rate signal.

Low gain gyro with lag around the switching amplifier for pseudo-rate signal.

These methods were compared on the analog computer and on an air bearing simulator using actual hardware. The best overall performance was obtained with the lead-lag network.

This network operated over the widest range of vehicle rates and always converged to limit cycle rates determined by single pulses which were less than the design rate of nine degrees per hour. This single pulse performance gives the theoretical minimum rate for an ON-OFF system and thus will minimize both photographic smear and gas usage.

A study was also made of the maneuver techniques that could be used. This comparison considered three techniques (two open loop and one closed loop): a high gain gyro with timed torquing impulse (open loop maneuver), a high gain gyro with capacitor feedback and timed torquing impulse (open loop maneuver), and the direct integration method described above for the Lunar Orbiter mission (closed loop maneuver). On the basis of reliability, weight, maneuver rate capability and accuracy, the direct integration of the vehicle rate method was chosen. In addition, direct verification that the vehicle is indeed moving can be ascertained from the telemetered rate output of the gyro.

The limit cycle operation of the vehicle in the presence of solar torques and gravity gradient has been studied on both the analog and digital computers. For the primary mission, there does not appear to be any appreciable effect on the limit cycle performance of the vehicle. However, during the extended mission, the wide angle excursion of the spacecraft -- coupled with the gravity gradient torques -- will probably cause a one-sided limit cycle to predominate. The solar pressure against the spacecraft is destabilizing and increases as a double angle function. As the limit cycle deadband is increased from 0.2 degree to 12.0 degrees, the unbalance torque from this solar pressure will increase approximately fifty times. This increase, coupled with the existing gravity gradient torque, will cause one-sided limit cycle operation -- but with approximately the same time between thruster firings as in the symmetrical limit cycle case.

II Subsystem Performance

The analysis of the attitude control performance data obtained from missions I, II, and III is by no means complete, so the results presented in this paper represent only highlights of the achievments (and problems) of the attitude control system. Where particularly pertinent, ground test data -- either component or spacecraft level -- are included for comparison with the space flight data. Because of some uniqueness of the data by mission phase, the data are grouped and presented approximately in mission chronology.

Atlas/Agena Launch

As mentioned in Section I, the ACS is essentially inactive during the launch phase so there are no flight performance data available. The most critical parameter is boost vibration survival. Ground tests at both component and spacecraft levels indicated that, in the ACS, the IRU and the Canopus Star Tracker (CST) were the most sensitive components. The IRU required several modifications following initial vibration testing due to broken part tie-downs and electrical leads on resistors and large capacitors. The CST required modifications due to sun shutter sticking and to shifts in its mechanical and optical roll error nulls.

In addition to the component modifications, the component sine vibration test levels were reduced for both the IRU and CST at certain critical frequencies as spacecraft tests proved that the component test levels were several factors higher than the inputs to the components during the spacecraft tests. The IRU random vibration levels were reduced at the same time. Figures 8, 9, and 10 show the IRU, CST, and spacecraft ground test vibration levels with maximum vibration levels measured during launch at the ACS location superimposed. It can be seen in Figures 8 and 9 that the maximum launch sine inputs to the IRU and CST correlate very well with the inputs during spacecraft ground tests. The random vibration inputs did not correlate as well, but the launch data are almost an order of magnitude lower than spacecraft ground test random vibration inputs to the IRU and CST.

There has been no evidence of any damage to, or performance degradation of, the IRU, LST, or any other ACS component during mission launch phases for L. O. I, II, or III.

Sun Acquisition and Re-acquisition

Each spacecraft initially acquired the Sun within the required one hour from launch; L.O. I within 53 minutes (10 minutes after spacecraft separation), L.O. II within 56 minutes, and L.O.III within 56 minutes.
Initial acquisition occured in all three missions

at a time when telemetry data were not available until either; after acquisition started, or; for L.O. II and III, after it was completed. Figure 11 shows the initial acquisition for L.O. I in Pitch from an angle of -27.1° (Sun Sensor telemetry saturation). This acquisition was accomplished with the Coarse Sun Sensor ON and in the narrow dead zone (approximately ±2.0° for Coarse Sun Sensor). The theoretical acquisition, with design tolerances is also shown.

A re-acquisition, performed during the extended mission, is also shown and is much closer to design tolerances. This plot is typical of the more than 250 re-acquisitions made for L.O. I during the mission, following Sun occultations (by the Moon) or following maneuvers away from the Sun for ΔV firings, photo sequences, or for thermal control. There were approximately 120 re-acquisitions for L.O. II, and 126 for L.O. III, using both narrow (±0.20°) and wide (±2.0°) dead zones with Coarse Sun Sensors.

Sun and Canopus Oriented Limit Cycle

As in Sun and Canopus acquisition, the sun sensors, CST, and subsystems performed within design tolerances during L.O. I and II limit cycle operation. Table 4 shows typical limit cycle values for Roll, Pitch, and Yaw axes of L.O. I and II. Since there was no absolute reference, these values represent Sun Sensor and CST telemetry readings. Figures 12a, b, and c show typical Roll, Pitch, and Yaw limit cycles for one orbit of L.O. II with the dead-zone set at ±2.0°. Figures 13a, b, and c show similar limit cycles with the dead-zones set at ±0.2°. These figures include both Inertial Hold limit cycles (S/C referenced to IRU RIM with the Sun and Canopus occulted) and Sun locked limit cycles.

The lack of symmetry of these Sun locked limit cycles -- particularly at sunrise -- shows the effect of a predominant disturbance, solar pressure.

Limit cycle rates were slightly higher than the theoretical minimums (approximately 2°/hr; predicted from 'one-shot' thruster ON times) but, except for sharp disturbances such as antenna rotation, were considerably lower than the design goal of 9°/hr and significantly less than the mission design requirement of 36°/hr (to minimize shutter-open photo smear).

Table 5 gives the range of limit cycle rates and the changes in limit cycle rate during thruster ON times for typical L.O. I and L.O. II orbits in the narrow (±0.2°) and wide (±2.0°) dead-zones.

	DEAD-ZONE DEG.	LIMIT CYCLE RATE MAX °/HR. 1)	LIMIT CYCLE RATE MIN °/HR	CHANGE IN LIMIT CYCLE RATE MAX °/HR. 1)	CHANGE IN LIMIT CYCLE RATE MIN °/HR
L.O. I					
Roll	±0.2	16	4	24	9
	±2.0	21	1.2	21	3.5
Pitch	±0.2	43	6	36	14
	±2.0	16	5	23	11
Yaw	±0.2	16	2.5	30	8
	±2.0	17	3	25	8
L.O. II					
Roll	±0.2	4.7	2.1	8.1	5.2
Pitch	±0.2	16	1.8	21	2.9
Yaw	±0.2	21	1.1	29	3.6

1) DISTURBANCES SUCH AS SUNRISE, ANTENNA POINTING, ETC.

Table 5 LIMIT CYCLE RATES

Inertial Hold Mode

The Inertial Hold Mode (IHM) of the ACS (IRU in RIM) was used during substantually longer portions of the L.O. missions than initially planned. This high usage was primarily due to the requirement to operate off-sun for thermal control and to the difficulties encountered with Roll Canopus Lock.

RIM (attitude) accuracy was obtained by comparing RIM outputs with Sun Sensor and Canopus Star Tracker outputs while the S/C was in IHM and the Sun and Canopus were in their respective sensor's field of view. The resolution of telemetry readings for the Fine Sun Sensor and CST was ±0.05°, and for the Coarse Sun Sensor was ±0.3°. The IRU, Sun Sensor and CST attitude readings all agreed within telemetry resolution errors when the IRU readings were corrected for drift and initial bias. The initial bias was caused by the IRU not being switched to RIM at a time when celestial sensors were at null (output = zero at time of switching).

SUN SENSOR MODE	DEAD-ZONE	DESIGN TOLERANCES DEG.	ROLL L.O. I DEG.	ROLL L.O. II DEG.	PITCH L.O. I DEG.	PITCH L.O. II DEG.	YAW L.O. I DEG.	YAW L.O. II DEG.
Fine	Narrow ±0.2°	±0.18±.003	+0.16 -0.18	±0.19	+0.193 -0.194	±0.20	+0.18 -0.18	±0.25
Fine	Wide ±2.0°	±2.0 ±0.3	NA	+2.00 -2.03	+1.87 -2.05	NOT USED	+1.92 -1.97	NOT USED
Occulted Fine Null			NA	NA	-0.05	+0.05	+0.04	+0.06
Occulted Coarse Null 1)			NA	NA	+0.07	+0.24	-0.79	+0.55

1) Telemetry resolution = ±0.3 degree. Values are averages of several readings.

Table 4 TYPICAL CELESTIAL-ORIENTED LIMIT CYCLES

IHM dead-zones were slightly narrower than the equivalent Sun- or Canopus-locked dead-zones. For example, Table 6 shows typical IHM dead-zones for L.O. I and II. Limit cycle rates in IHM were very similar to those shown in Table 5 for Sun- and Canopus-locked limit cycles.

		DEADZONE	
L.O. I	Axis	Narrow(±0.20°)	Wide(±2.0°)
	Roll	+0.16 -0.18	1) NA
	Pitch	+0.17 -0.18	1) NA
	Yaw	+0.16 -0.18	1) NA
L.O. II	Roll	±0.167	±1.98
	Pitch	+0.171 -0.166	+1.97 -1.99
	Yaw	+0.162 -0.186	+1.94 -1.97

1) Wide deadzone used very little during L.O. I mission.

Table 6 IHM DEADZONES

ΔV Maneuvers

ΔV maneuvers were performed during translunar trajectory (Midcourse Maneuver), Moon arrival (Orbit Injection), and during orbit around the Moon (Orbit Transfer).

There are several errors, such as thrust pointing error, accelerometer error, acceleration integration error and time error, which contribute to the total ΔV maneuver error. These individual error contributors have not been identified separately in the flight data, but the total ΔV maneuver errors have been well within the specification allowance. Figures 14, 15 and 16 illustrate the effect of the small ΔV maneuver errors by comparing the desired trajectories and orbits with those actually achieved by L.O. I, II, and III.

Photo Maneuvers

There were eleven 3-axis photo maneuvers during the L.O. I mission from which maneuver accuracies for Roll, Pitch and Yaw could be determined. The pre-photo maneuvers and the total 3-axis errors (computed) in those maneuvers, and the range of attitude errors and rates during the photo time are shown in Table 7.

The limit cycle attitudes at the initiation of the Roll, Pitch, and Yaw maneuvers were all within the ±0.2° deadband requirements except for site A-7 where a roll update was not performed just prior to the start of the Roll maneuver. Attitude errors during photo time were all less than the ±0.2° design requirement (IHM narrow dead-zone), and attitude rates were all less than the ±56°/hr photo requirement. All but two, of approximately 12°/hr, were less than the design goal of ±9°/hr.

Similar data is given for L. O. II in Table 8. The maneuver errors were somewhat less than those for L. O. I, primarily because of better Canopus updating and a better technique for starting the maneuver which minimized the Roll errors. This technique is discussed under CST problems. Maximum rates during photo times were somewhat higher than for L. O. I, though still less than 2/3 of the photo requirement.

Attitude maneuver rates are given in Table 9. It can be seen that the minimum rates are below the design requirement and the maximum rates are all within the design requirement. Since the V/F converter actually maintained its accuracy over a wider range of input rates (voltages) than design requirements specified, the errors in maneuver rates did not contribute significantly to maneuver errors.

Figure 17 shows typical IHM limit cycles during a photographic sequence with L. O. II. The sequence starts with Roll, Pitch and Yaw gyro channels in narrow (±0.2°) dead-zone RIM. The Roll channel switches to RM during Roll maneuver and then back to RIM at completion of Roll maneuver. This is followed by similar maneuvers about the Pitch and Yaw axes.

Several maneuver accuracy tests were performed in a manner which allowed Sun Sensor and CST outputs to be used for calibration of the maneuver angles. This was accomplished by starting a 360° maneuver with the Sun or Canopus in the sensor FOV. The difference between the sensor readings at the finish of

	ATTITUDE MANEUVER - DEG.			SPACECRAFT ATTITUDE CHANGE DURING PHOTO - DEG.				ATTITUDE RATE DURING PHOTO - DEG/SEC		
SITE	ROLL	YAW	PITCH	ROLL	YAW	PITCH	CRAB ANGLE	ROLL	YAW	PITCH
A-0	+3.60	+12.45	-8.10	+.04 to +.13	0 to +.13	-.13 to +.14	0 to +.9	+.0003	-.0018	+.0018
A-2	+5.43	+12.48	-4.98	-.10 to -.15	+.10 to +.14	-.06 to -.12	+.2 to +.5	+.0083	-.0035	+.0015
A-3	+5.70	+12.60	-1.60	-.06 to -.12	+.05 to +.10	-.02 to +.04	+.15 to.17	+.0015	-.0025	+.0015
A-4	+5.90	+12.50	-2.40	-.07 to -.11	+.05 to +.14	-.09 to +.03	+.1 to +.3	+.0010	-.0024	+.0019
A-5	+6.50	+12.60	-2.30	-.06 to -.10	+.16 to +.11	-.05 to -.09	+.1 to+.15	+.0011	-.0033	+.0012
A-6	+6.90	+12.90	-16.7	+.12 to +.14	-.03 to +.07	+.06 to +.12	+.1 to +.2	-.0002	-.0017	+.0020
A-7	+7.60	+12.80	-12.9	0 to +.06	+.01 to -.08	0 to +.09	+1.0	+.0013	-.0018	+.0020
A-8.1	+8.04	+12.81	-11.3	+.02 to +.15	+.03 to -.18	-.06 to -.15	0 to +.15	+.0021	+.0024	+.0008
A-9.2a	+8.11	+12.81	-6.86	+.12 to +.14	+.05 to +.08	-.03 to +.05	+.15	0	-.0010	+.0012
A-9.2b	+8.20	+12.80	-8.30	+.07 to +.13	+.16	+.05 to +.15	+.15 to+.5	+.0017	0	+.0022

Table 7 ATTITUDE MANEUVERS - L.O. I

PHOTO SITE	ATTITUDE MANEUVER DEGREES			MAXIMUM RATE DURING PHOTO DEGREES/SECOND			ATTITUDE CHANGE DURING PHOTO DEGREES			CRAB ANGLE VARIATION
	ROLL	PITCH	YAW	ROLL	PITCH	YAW	ROLL	PITCH	YAW	DEGREE
P-1	+5.24	+2.81	+9.25	+.00239	+.00260	-.00240	-.17 to -.07	-.02 to +.15	+.11 to +.70	-.05 to +.18
S-1	+5.24	+2.81	+9.25	+.00239	+.00260	-.00240	-.17 to -.07	-.02 to +.15	+.11 to +.70	-.05 to +.18
S-2a	-1.47	+11.12	+6.93	0	+.00403	-.00390	+.145	+.14 to +.10	+.09 to +.06	+.03 to +.20
S-2b	+12.17	-8.40	+11.70	+.00213	-.00391	-.00196	-.16 to -.14	+.06 to -.02	+.16 to +.13	-.30 to 0
P-2	+5.30	-2.40	-2.40	-.00217	+.00300	-.00467	-.08 to -.09	+.14 to +.15	+.05 to +.04	+.15 to +.20
P-3a	+5.33	+5.45	+9.50	+.00204	+.00382	-.00226	-.12 to -.09	-.05 to +.03	+.14 to +.12	-.18 to -.15
P-3b	+4.25	+5.33	+9.56	+.00304	+.00208	-.00200	-.15 to -.10	+.02 to +.05	+.07 to +.05	-.05 to +.10
P-4	+5.33	+5.45	+9.50	+.00117	+.00304	-.00652	-.16 to -.15	+.02 to +.07	+.08 to +.07	0.00
P-5	+8.66	-1.25	+10.82	-.00209	-.00221	-.00179	-.12 to -.14	-.08 to -.12	+.14 to +.09	-.15 to +.10
P-6a	+5.38	-7.74	+9.62	+.00167	0	-.00067	-.08 to -.06	+.155	+.06 to +.05	+.13 to +.20
P-6b	+5.39	-9.23	+9.65	+.00291	0	-.00305	-.13 to -.11	+.140	+.01 to -.04	+.10 to +.15
S-6	+16.77	+13.20	+8.30	+.00226	+.00087	+.00104	+.16	+.15	-.05	-.12
S-7	-59.19	+5.91	+9.72	0	+.00107	-.00100	-.14	-.09	+.14	+.208
S-8	+5.40	-8.52	+9.82	+.00127	0	-.00670	+.16 to +.14	+.05	+.16 to +.15	+.07 to +.05
S-9	+5.42	-0.63	+9.87	-.00100	+.00017	+.00200	-.07	+.15	+.16	+.025
P-7a	+5.47	0	+9.79	-.00230	+.00267	-.00280	-.14 to -.17	+.06 to +.10	+.03 to 0.00	-.05 to +.05
P-7b	+5.47	-1.50	+9.82	+.00252	0	-.00305	-.14 to -.12	+.15	+.04 to 0.00	+.10 to +.15
S-10	+5.42	+4.14	+9.93	+.00087	+.00230	-.00113	-.12	-.10	+.065	-.075
P-8a	+5.43	-7.20	+9.97	+.00300	0	-.00100	-.17 to -.12	+.10	+.12 to +.14	-.15 to +.15
P-8b	+5.44	-8.69	+9.95	+.00100	-.00200	+.00190	-.14 to -.11	+.06 to +.02	+.06 to +.08	-.20 to -.10
P-8o	+5.47	-10.17	+9.91	+.00113	-.00233	-.00117	-.14 to -.10	+.16 to +.11	+.16 to +.14	-.07 to +.10
S-11	+5.50	+11.49	+9.95	-.00100	+.00368	+.00433	-.17 to -.15	-.14 to -.10	+.14 to +.16	-.17 to -.30
P-9	+5.50	-4.20	+10.00	-.00220	+.00100	-.00090	-.06 to -.01	+.14 to +.15	+.12 to +.09	-.07 to +.20
P-10a	+5.49	+6.82	+9.98	+.00060	+.00213	-.00033	-.13 to +.12	0.00 to +.05	+.10 to +.09	0.00 to +.20
P-10b	+5.50	+5.33	+10.01	0	0	-.00087	-.14	+.15	+.04 to +.02	-.25 to -.05
S-12	+73.09	+9.00	+25.06	+.00200	-.00100	+.00043	-.05	+.05	-.17	+.208
P-11a	+5.55	-7.44	+10.00	+.00078	-.00190	-.00193	-.16	-.06 to -.12	+.14 to +.10	-.20 to -.05
P-11b	+5.57	-8.92	+10.02	+.00118	-.00190	-.00080	-.16 to -.14	+.11 to +.08	+.13 to +.12	-.05 to 0.00
P-12a	+5.55	+2.66	+10.08	+.00160	+.00173	+.00487	-.14 to -.11	+.14 to +.16	-.04 to +.06	-.15 to -.10
P-12b	+5.57	+1.17	+10.11	+.00133	+.00300	-.00170	-.14 to -.12	-.02 to +.08	+.10 to +.08	-.10 to +.10
S-13	+5.52	+6.16	+10.23	-.00113	+.00195	-.00200	-.17	-.10	+.14	-.40
S-14	-178.00	0.00	0.00	0	+.00100	-.00152	+.06	---	---	+.208
P-13a	+5.62	+1.62	+10.16	+.00073	+.00230	-.00117	-.11 to -.09	+.07 to +.12	+.07 to +.05	+.05 to +.20
P-13b	+5.60	+0.12	+10.20	-.00077	+.00317	-.00130	-.15 to -.16	+.09 to +.15	+.10 to +.08	0.00
S-15	+12.60*	0.00	+29.70	0	+.00090	+.00370	-.14	-.08	-.16	+.208
S-16	+5.52	-7.67	-10.30	+.00080	+.00207	-.00170	-.14	-.05	+.04	-.35
S-17	+63.40	+25.00	+15.20	0	+.00217	+.00240	+.18	+.04	-.17	+.208

* +60.1 AFTER YAW

Table 8 ATTITUDE MANEUVERS - L. O. II

AXIS	ROLL				PITCH				YAW			
DEAD-ZONE	NARROW		WIDE		NARROW		WIDE		NARROW		WIDE	
Req't.~°/sec	.495 to .605		.050 to 1.05		.495 to .605		.050 to 1.05		.495 to .605		.050 to 1.05	
L. O. I												
Max.Rate°/sec	In Spec		0.051		In Spec		0.059		TBA		TBA	
Min.Rate°/sec	In Spec		0.047		In Spec		0.036		TBA		TBA	
L. O. II	Plus	Minus	Plus	Minus	Plus	Minus	Plus	Minus	Plus	Minus	Plus	Minus
Max.Rate°/sec	.503	.521	.058	.052	.521	.521	.066	.054	.512	.512	NA	.052
Min.Rate°/sec	.486	.484	.058	.040	.494	.494	.048	.045	.494	.487	NA	.046

Table 9 ATTITUDE MANEUVER RATES

	LUNAR ORBITER I			LUNAR ORBITER II			LUNAR ORBITER III					
							Roll		Pitch		Yaw	
	Roll	Pitch	Yaw	Roll	Pitch	Yaw	NDZ	WDZ	NDZ	WDZ	NDZ	WDZ
Star Map, to Canopus and Other	32	5	8	11	6	0	12	2	6	0	2	0
Thermal Pitch-off	0	77	0	0	9	0	0	0	68	10	0	0
Attitude Update	77	29	45	14	3	13	4	5	2	0	1	0
Photo	40	28	24	80	66	70	104	0	94	0	82	0
Velocity	8	8	0	6	6	0	6	0	6	0	0	0
Sub-Total	150	147	77	111	90	83	126	7	176	10	85	0
TOTAL		374			284				404			
Number in Narrow Dead-zone	149	147	77	98	82	78		126		176		85
Number in Wide Dead-zone	1	0	0	13	8	5		7		10		0
Photo Maneuvers 3-axis		12			34					41		
2-axis		2			1					8		
1-axis		6			4					1		
	Dead-zone			Dead-zone			Dead-zone					
Celestial Acquisitions;	Narrow	Wide	TOTAL	Narrow	Wide	TOTAL	N_{a}rrow		Wide		TOTAL	
Canopus	4	0	4	139	7	146	152		19		171	
Sun	143	1	144	99	21	120	121		5		126	

Table 10 ATTITUDE MANEUVER SUMMARY - PHOTO MISSION ONLY

the maneuvers and those at the start were the maneuver errors. These errors are summarized for L. O. I, II and III in Figure 18.

A summary of all photo mission maneuvers performed by L. O. I, II and III is given in Table 10.

Reaction Jet Gas Usage

The cold-gas (nitrogen) usage of the ACS reaction control systems of L. O. I, II and III has been in excess of that predicted for these missions. This was not caused by excessive usage for the planned events, however, but by the large number of unplanned events - primarily maneuvers - required for solving in-flight problems and obtaining data for plotting successive missions. Comparisons of predicted vs actual gas usage for L. O. I, II and III are shown in Figure 19. The bases for the curves are:

a. "Minimum Estimate" and "Conservative Estimate" predicted for planned events.
b. "Minimum Actuals" and "Conservative Actuals" re-calculated for actual events using the same prediction usage coefficients as for a.
c. "Actual" usage based on P V T telemetry measurements of gas supply during mission.
d. "Actual" usage based on a "Dynamic" computer program using telemetry measurements of changes in angular rates, moments of inertia calculated from use of expendables, and estimates of jet thrust, moment arms and specific impulse.

The nitrogen prediction parameters were:

Limit Cycle	
Thrust	0.062 lb
ISP	68 sec
Pressure	23 psia
Minimum Pulse width	12 msec
Residual Rate	1.8 to 36 deg/hr
Maneuvers	
Thrust	0.056 lb
ISP	71 sec
Pressure	19.5 psia
Maneuver Rate	0.55 deg/sec

It can be seen that both "actuals" fall within or slightly below the range of gas usage given by the "Minimum Actual" to "Conservative Actual" computations per b. Because actual events were closer to planned events for L. O. II, the total range of gas usage between planned and actual was smaller.

It is felt that either the P V T or "Dynamic" predictions of gas usage represent acceptable accuracy for mission use when events come off as planned. Problems with L. O. in Canopus tracking and S/C thermal balance caused the greatest deviations from planned events for L. O. I, II and III.

IRU Drift vs. Life

During component level testing of IRU's, the g-sensitive and g-insensitive gyro drift was a critical parameter because it was a major problem with IRU yield. During the space flights of L. O. I II and III it again was a critical parameter because the spacecraft thermal balance problem and the CST glint problem required that the ACS be in Roll, Pitch and Yaw IHM for large percentages of the mission times. Where drift was large, as for L.O. I Yaw, this meant many maneuvers to up-date IHM null with celestial information. As mentioned above, these maneuvers had significant impact on the ACS nitrogen usage. L. O. III Pitch, Roll, and Yaw gyros had particularly low g-insensitive drift rates which were very stable.

Figure 20 shows the space flight g-insensitive drift histories of the Roll, Pitch and Yaw gyros from L. O. I, II and III. These drift data were obtained from simultaneous telemetry measurements of IRU RIM outputs and Sun Sensor outputs vs time.

Figure 21 shows the total ground test and space flight drift history for the L. O. III IRU through the first 30 days of the mission.

Table 11 gives the operational life of the IRU's for L. O. I, II and III through May, 1967, during which time only one temporary performance discrepancy occured. The Roll gyro in L. O. II IRU hung up several times in RIM, but was freed each time that it did.

		IRU START/STOP		GYRO					
		HRS		P/N	HRS	P/N	HRS	P/N	HRS
L. O. I (S/N 106 - MAINSTREAM)[1]	GROUND TEST	864	183	-47	703	-31	921	-33	886
To 11/1/66	SPACE FLIGHT	2000			2000		2000		2000
	TOTAL	2864			2703		2921		2886
L. O. II (S/N 109 - MAINSTREAM)[2]	GROUND TEST	1574		-22	1630	-23	1634	-24	1602
To 4/20/67	SPACE FLIGHT	3800			3800		3800		3800
	TOTAL	5374			5430		5434		5402
L. O. III (S/N 113 - BACKUP)[2]	GROUND TEST	392		B11	417	B10	417	B 4	424
To 4/20/67	SPACE FLIGHT	1800			1800		1800		1800
	TOTAL	2192			2217		2217		2224

1) DELIBERATELY CRASHED ON MOON
2) STILL OPERATIONAL

Table 11 I R U OPERATIONAL LIFE

C S T Glint

A very serious problem with Canopus acquisition and tracking occured during most of the L. O. I mission. Glint from the CST baffles - illuminated by reflections from the spacecraft, primarily the omni-antenna and the edges of the solar panel - prevented normal Canopus acquisition and tracking.

Figure 22 shows the star map output of the CST during a 360° star map Canopus acquisition maneuver. Superimposed is the ***a priori*** star map, adjusted after the fact, to the same initial Roll orientation. As can be seen, Canopus was not detected at all - although the Moon was detected. The Moon was used as the Roll reference for the midcourse ΔV maneuver when the Canopus acquisition was not successful.

It was determined that if, after midcourse, CST was turned first OFF and then ON while Canopus was in the FOV, it would acquire and track Canopus. This technique was used during the remainder of the L. O. I mission with the high-gain antenna pattern relative to Earth orientation being used to accurately point the CST at Canopus.

Later in the L. O. I mission an experiment was performed to verify the hypothesis that reflected sunlight from the omni-antenna (just outside the CST FOV) was the source of the baffle glint. This experiment oriented the spacecraft as shown in Figure 23 and then rolled it 360°. This maneuver accomplished three objectives;

1. It shaded the omni-antenna from the Sun behind the spacecraft while the CST looked at deep space (Position 2 of Figure 23).
2. It checked proper position and operation of the Sun shutter by having the CST look directly at the Sun (Positions 4 and 5).
3. It confirmed the glint theory by having the CST look at Canopus and at deep space with the omni-antenna illuminated by the Sun (Positions 1 and 4).

The star map output during this experiment is also given in Figure 23 and was tentatively interpreted as verifying the glint phenomenon, assessing the omni-antenna as the primary offender.

Subsequent to the L. O. I mission, but prior to L. O. II, a ground glint test was run on a spacecraft within a dark room and with a MgO plate simulating the CST aperture.

A photometer measured the glint incident on this plate as various small areas (less than 50 in^2 each) of the spacecraft were illuminated from a simulated Sun source. The total spacecraft glint was calculated relative to the illumination intensity into the CST aperture necessary to cause it to recognize and track the source. (Interference level equal to 1.0) The results of this test on an unmodified spacecraft are given in Table 12.

SPACECRAFT #1 NO MODIFICATIONS	Interference Level
1. Omni Antenna and Boom, White Paint	1.090
2. Solar Panel #2, Edge Unpainted	0.068
3. Solar Panel #4, Edge Unpainted	0.021
4. High Gain Antenna (Multiple Refl.)	0.035
5. EMD (Multiple Reflection), Solar Panel Top - Flat Black, Edge Unpainted	0.400
TOTAL	1.614
SPACECRAFT #1 MODIFIED TO REDUCE INTERFERENCE	
1. Omni Antenna Boom - Gloss Black; Discs - Flat Black	0.027
2. Solar Panels #2 & #4, Edges Taped Black -- Below Noise Level	-
3. High Gain Antenna -- Below Noise Level	-
4. EMD (Multiple Reflections) Solar Panel Top - Flat Black, Edges Taped Black	0.210
TOTAL	0.237

Table 12 RESULTS OF GROUND GLINT TESTS

After modification as identified in the lower portion of the Table, the interference levels were reduced significantly as shown.

The spacecraft L. O. II and III were modified as indicated above prior to launch, and the Canopus acquisition, re-acquisition and tracking data were normal. Figure 24 shows the ***a priori*** and the actual initial acquisition star map for L. O. II.

Spaceflight A C S Performance Summary

All mission objectives of the A C S were met within system specifications including;

1. Translunar celestial and inertial orientation.
2. Attitude maneuvering and inertial attitude hold for ΔV maneuvers.
3. Orbital celestial and inertial orientation.
4. Attitude maneuvering and inertial attitude hold for photo sequences.

All A C S components performed within design requirements except the CST which suffered from a glint problem. This was, to a large extent, eliminated after the first mission by repainting white portions of the spacecraft black.

The closed-loop technique for attitude maneuvering (slewing) of the spacecraft proved to be very accurate and highly flexible, operationally -- the latter attribute being very important during the solution of mission problems such as thermal balance and Canopus glint.

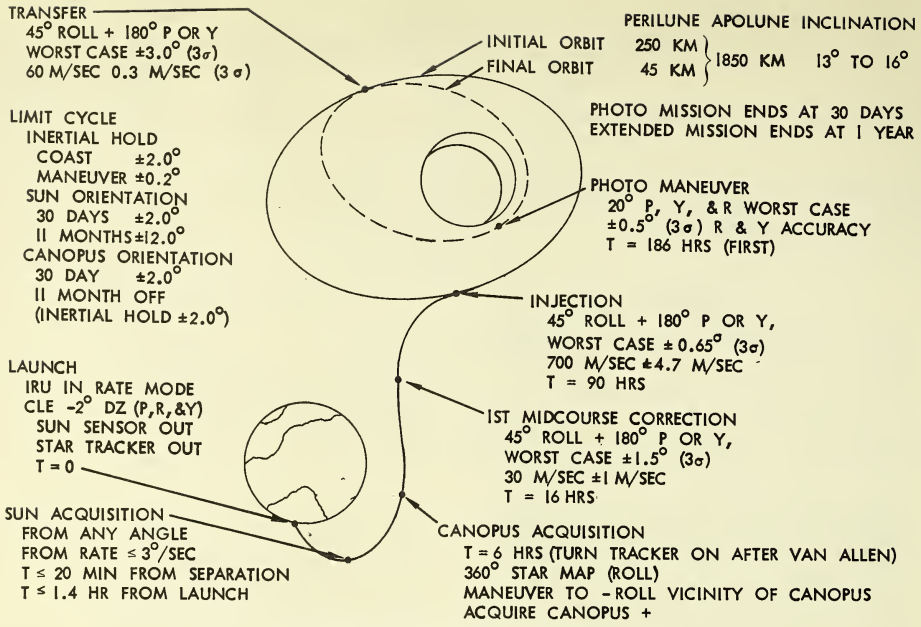


FIGURE 1 ACS REQUIREMENTS

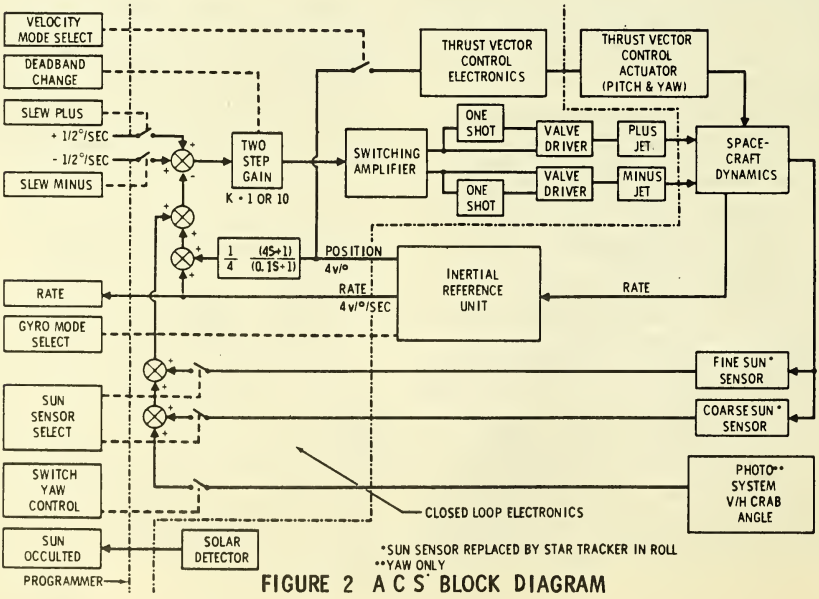


FIGURE 2 A C S BLOCK DIAGRAM

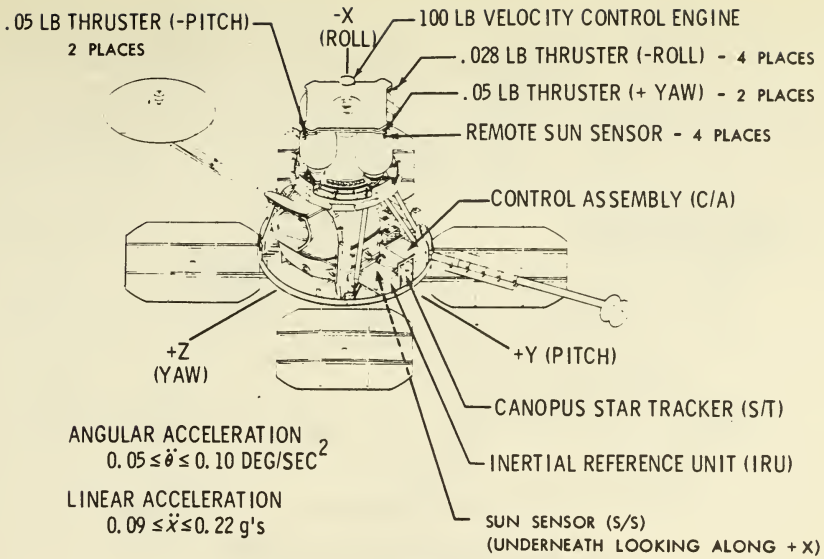


FIGURE 3 L. O. ACS COMPONENTS LAYOUT

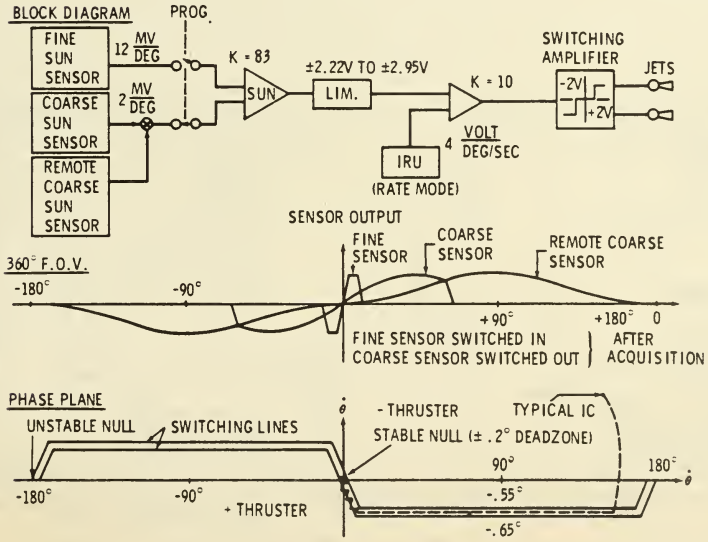


FIGURE 4 SUN ACQUISITION

67-533

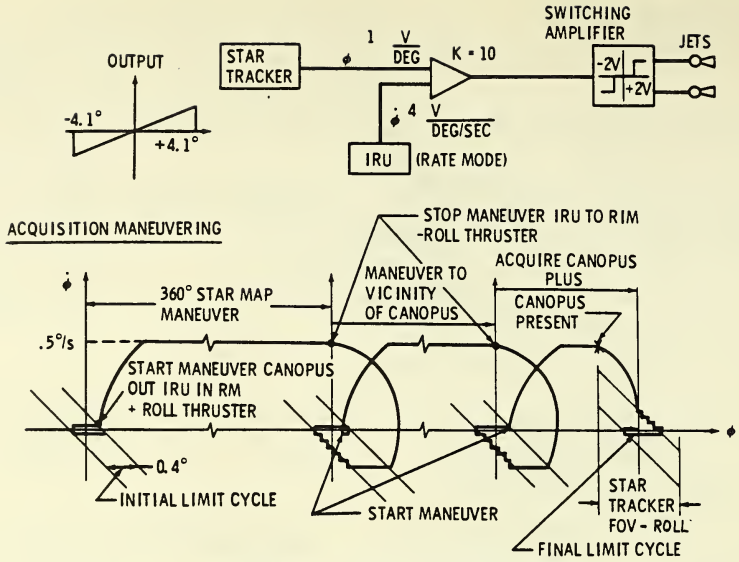


FIGURE 5 CANOPUS ACQUISITION

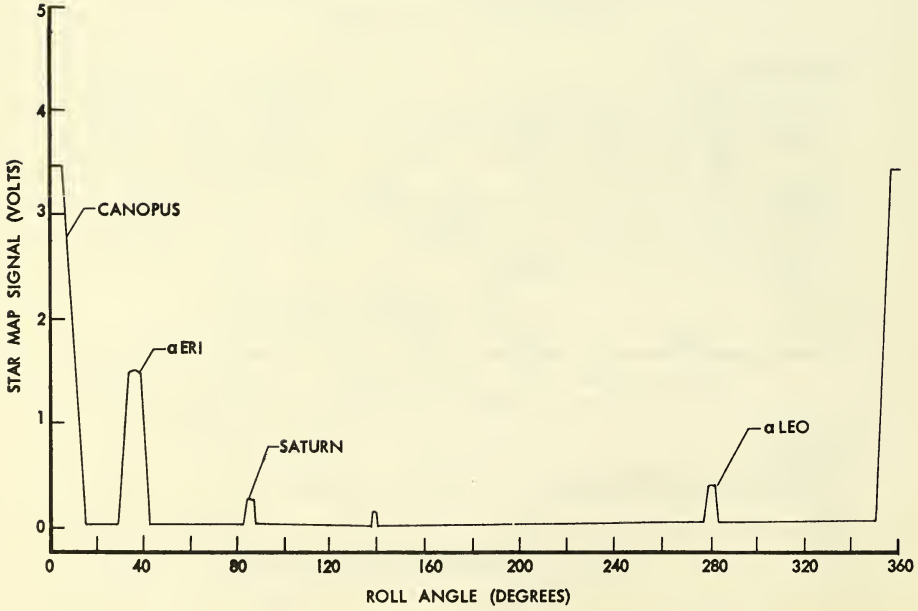


FIGURE 6 STAR MAP PREDICTION

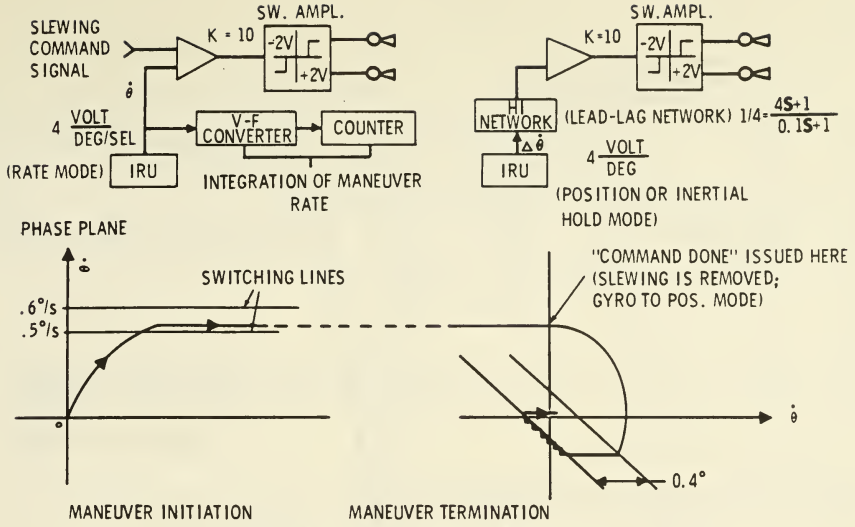


FIGURE 7 ATTITUDE MANEUVERING - ΔV AND PHOTO MANEUVERS

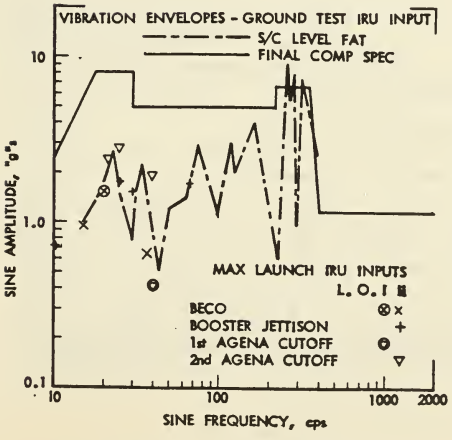


FIGURE 8 I R U SINE VIBRATION ROLL (X) AXIS

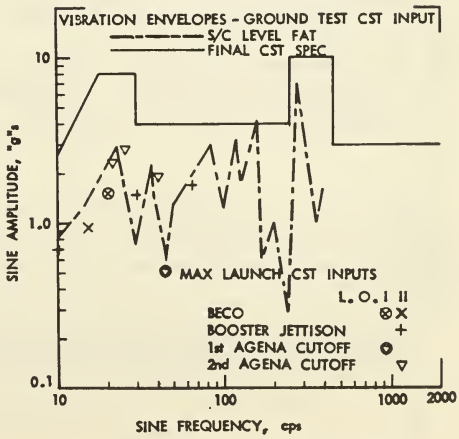


FIGURE 9 C S T SINE VIBRATION ROLL (X) AXIS

67-533

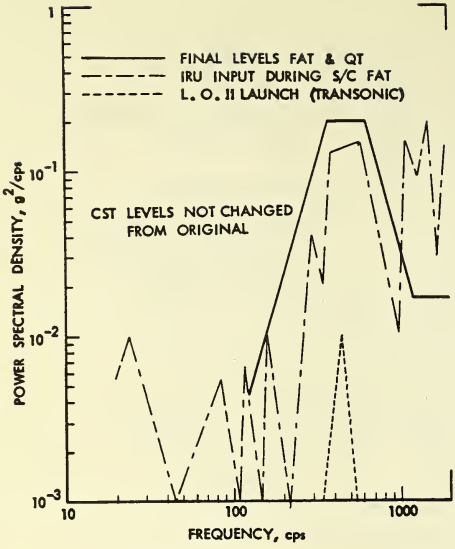


FIGURE 10 C S T AND I R U RANDOM VIBRATION

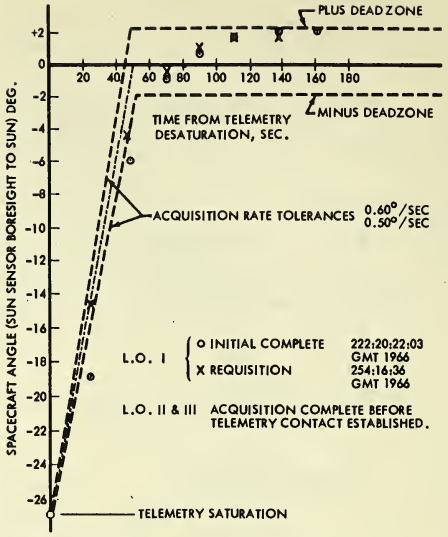


Figure 11: SUN ACQUISITION - PITCH

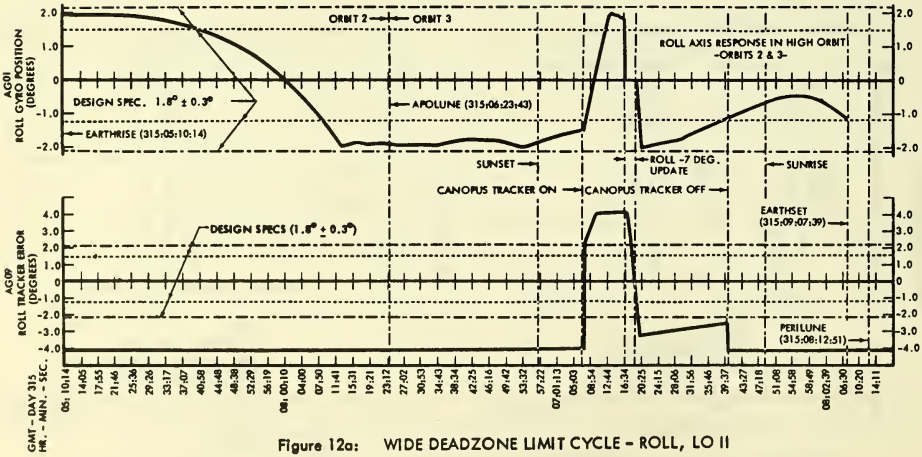


Figure 12a: WIDE DEADZONE LIMIT CYCLE - ROLL, LO II

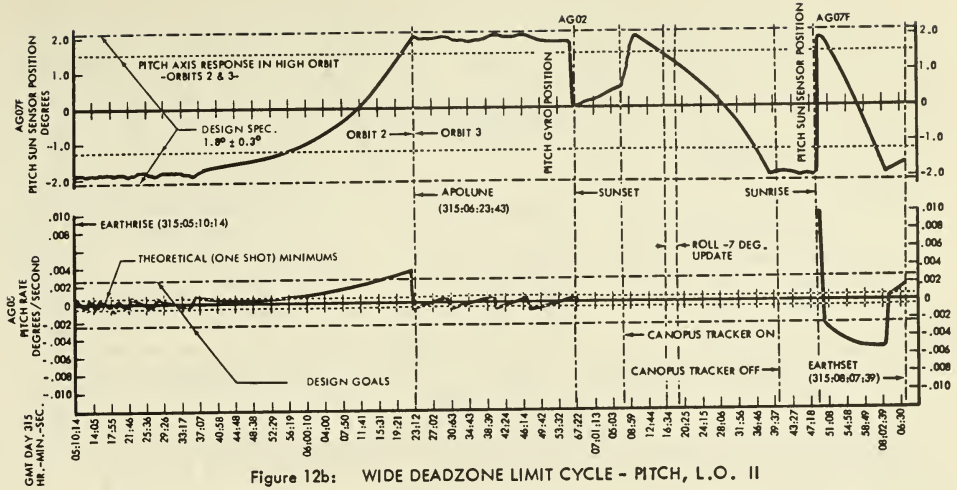


Figure 12b: WIDE DEADZONE LIMIT CYCLE - PITCH, L.O. II

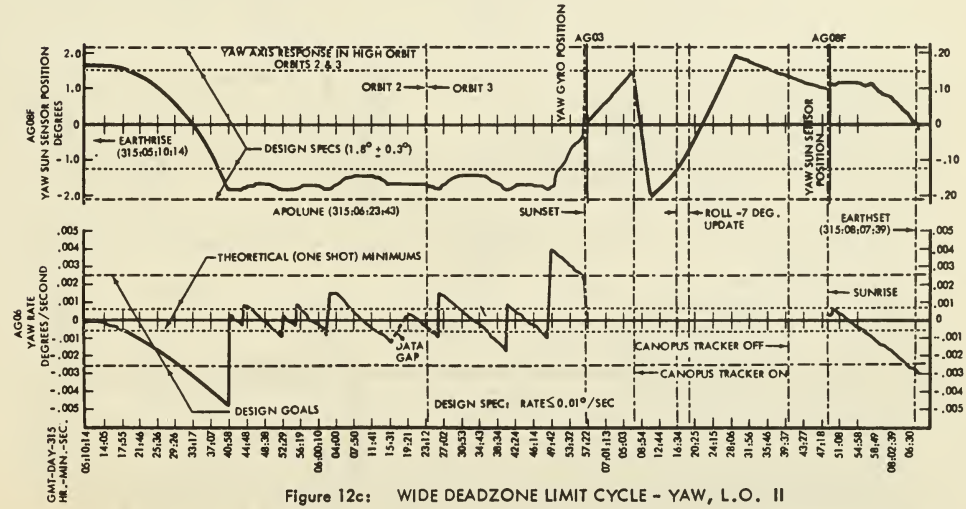


Figure 12c: WIDE DEADZONE LIMIT CYCLE - YAW, L.O. II

67-533

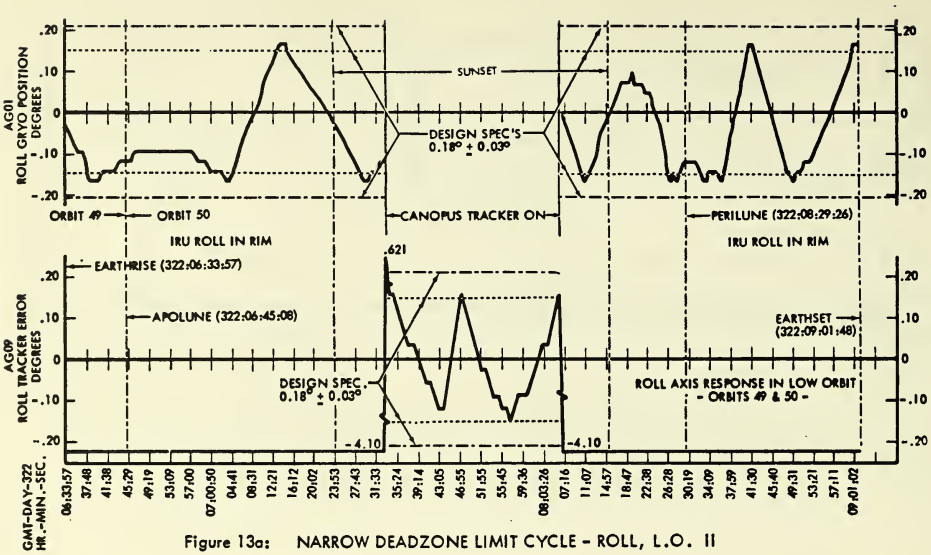


Figure 13a: NARROW DEADZONE LIMIT CYCLE - ROLL, L.O. II

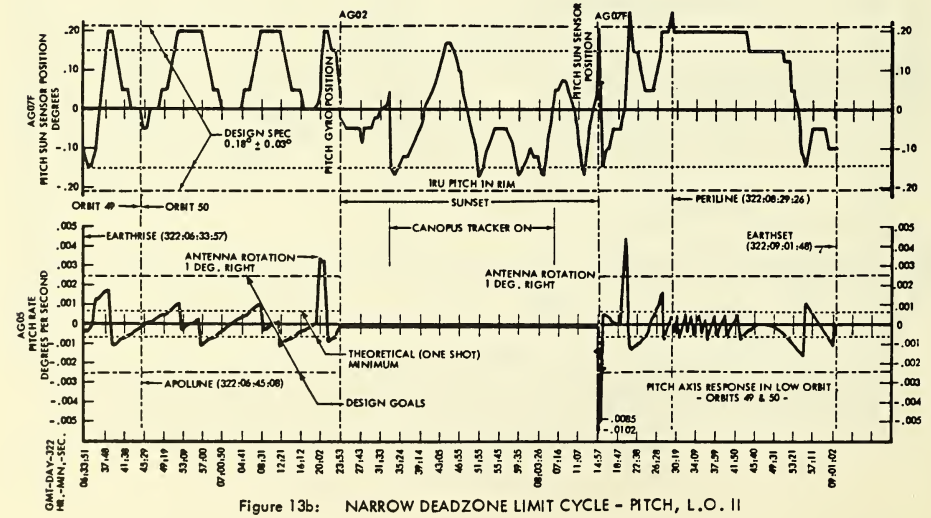


Figure 13b: NARROW DEADZONE LIMIT CYCLE - PITCH, L.O. II

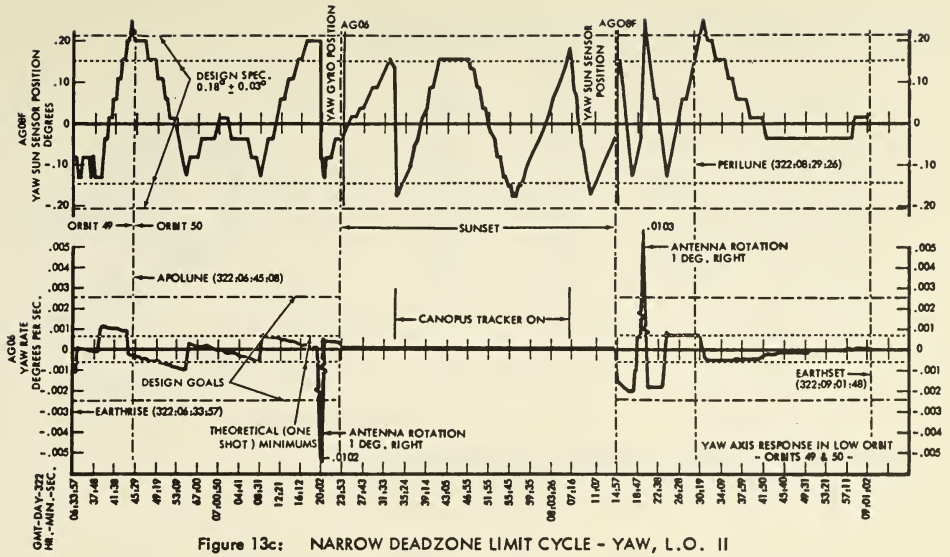


Figure 13c: NARROW DEADZONE LIMIT CYCLE - YAW, L.O. II

MISSION NUMBER	DESIRED TRAJECTORY (KM TO MOON CENTER)	ACTUAL TRAJECTORY (KM TO MOON CENTER)
I	6509	655
II	6023	6056
III	6123	6141

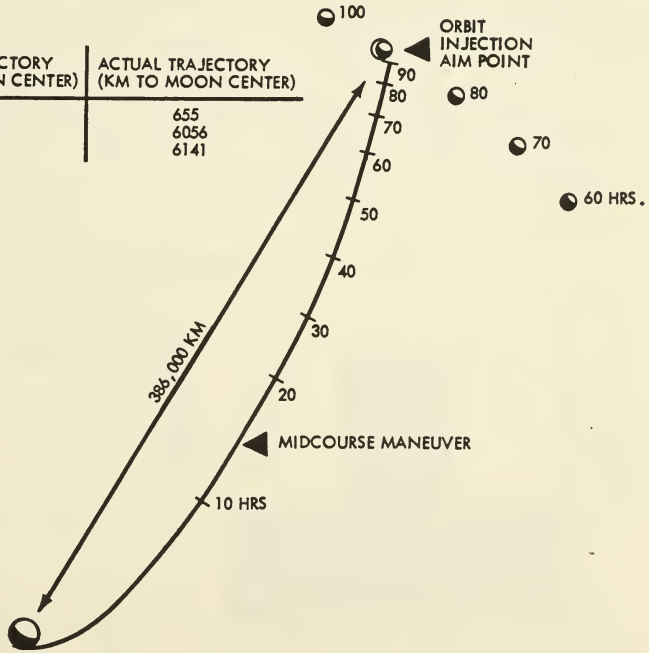


Figure 14: LUNAR ORBITER MIDCOURSE MANEUVER SUMMARY

67-533

MISSION NUMBER	ORBIT ALTITUDE	DESIRED TRAJECTORY (KM TO MOON SURFACE)	ACTUAL TRAJECTORY (KM TO MOON SURFACE)
I	H_p	199	189
	H_a	1850	1867
II	H_p	202	196
	H_a	1850	1867
III	H_p	213	210
	H_a	1850	1802

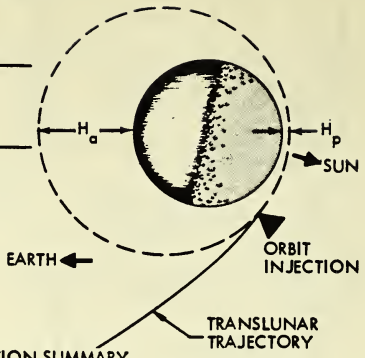


Figure 15: LUNAR ORBITER ORBIT INJECTION SUMMARY

MISSION NUMBER	ORBIT ALTITUDE	DESIRED TRAJECTORY (KM TO MOON SURFACE)	ACTUAL TRAJECTORY (KM TO MOON SURFACE)
I	H_p	57.9	56
	H_a	1855	1853
II	H_p	50.2	49.7
	H_a	1858	1853
III	H_p	54.8	54.9
	H_a	1846.4	1847.4

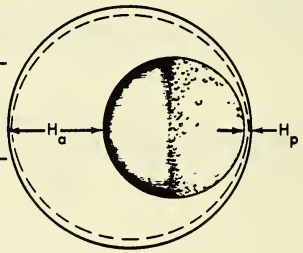


Figure 16: LUNAR ORBITER ORBIT TRANSFER SUMMARY

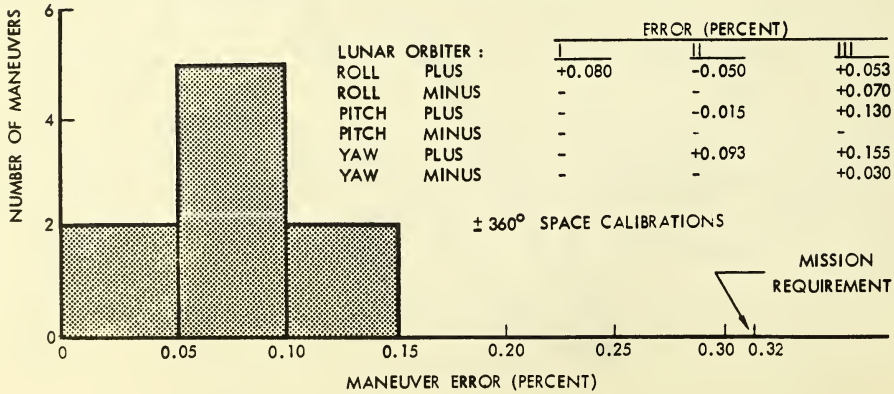


Figure 18: ATTITUDE MANEUVER ERROR

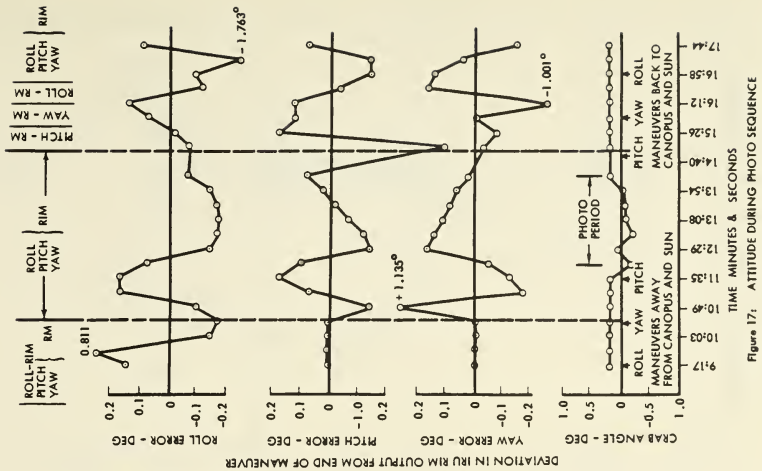


Figure 17: ATTITUDE DURING PHOTO SEQUENCE

SIGNIFICANT FACTORS

L.O.I
- STAR MAPS & ROLL UPDATE
- THERMAL MANEUVERS
- VCS REGULATOR LEAKAGE
- HIGH GYRO DRIFT
- PHOTO MANEUVERS

L.O. II
- IMPROVED OPERATIONAL PROCEDURES
- NOMINAL S C PERFORMANCE
- PHOTO MANEUVERS

L.O. III
- IMPROVED OPERATIONAL PROCEDURES
- LOW GYRO DRIFT
- INCREASED THERMAL MANEUVERS
- PHOTO MANEUVERS

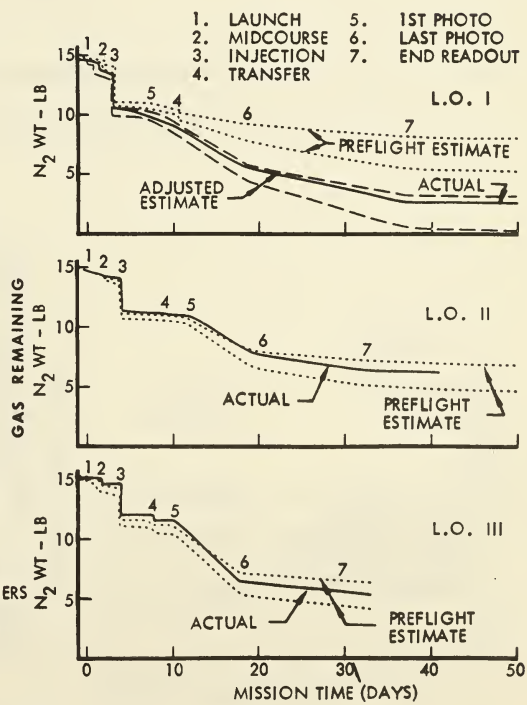


Figure 19: MISSION NITROGEN USAGE

67-533

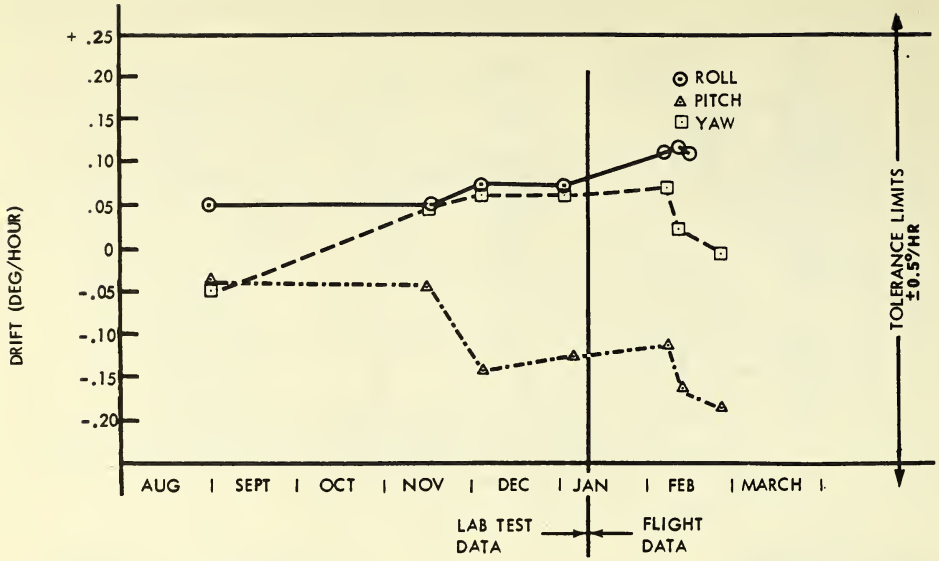


Figure 21: IRU FIXED TORQUE DRIFT LUNAR ORBITER III

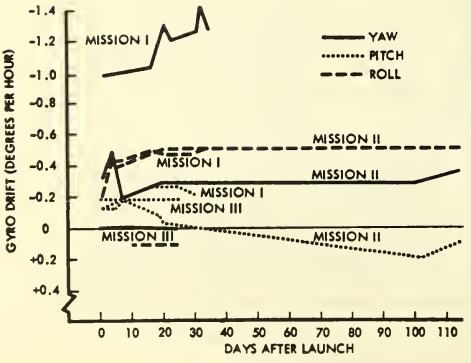


Figure 20: IRU DRIFT

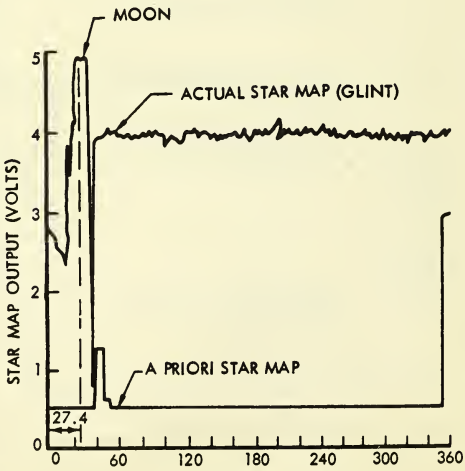


Figure 22: CANOPUS STAR TRACKER STAR MAP — LO I GLINT PROBLEM

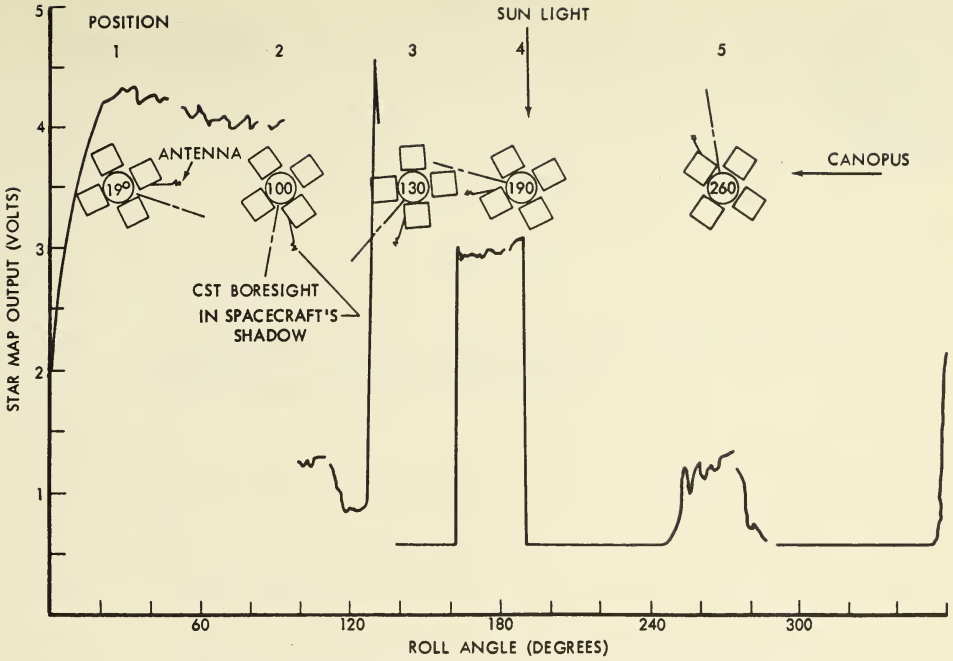


Figure 23: CANOPUS STAR TRACKER, LUNAR ORBITER I SPACE GLINT TEST

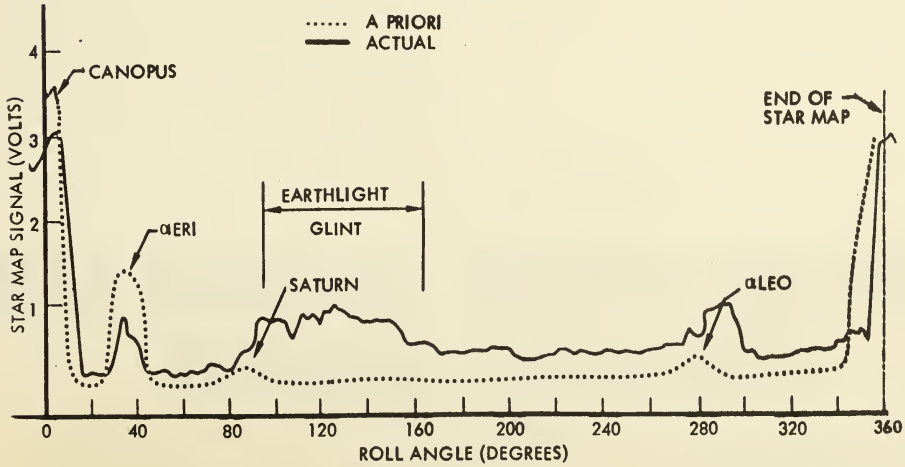


Figure 24: STAR MAP - LO II

67-533

-- NOTES --

No. 67-534

ATTITUDE CONTROL AND PRECISION POINTING OF THE APOLLO TELESCOPE MOUNT

by

W. B. CHUBB, D. N. SCHULTZ,
and
S. M. SELTZER
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-534

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

7.01, 7.03, 7.04

ATTITUDE CONTROL AND PRECISION POINTING OF THE APOLLO TELESCOPE MOUNT

W. B. Chubb, D. N. Schultz, and S. M. Seltzer
Guidance and Control Systems Analysis Branch
Guidance and Control Division, Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

The prime requirement of the Apollo Telescope Mount is to accurately point an experiment package, containing a complement of solar experiments, to acquire data on solar phenomena. A vehicle configuration with an Apollo Telescope Mount will be placed in a 485-kilometer circular orbit to give the experimenter an essentially atmosphere-free environment for solar observation. Experiment requirements dictated the development of a vehicle attitude control system, an experiment pointing system, and a roll positioning mechanism. These systems provide the capability of acquisition, offset pointing, automatic attitude hold operation, and manual operation. The vehicle attitude control system uses three two-degree-of-freedom control moment gyroscopes to supply control torques about three orthogonal vehicle axes. The experiment pointing system consists of a two-axis gimbal trim system to provide pitch and yaw positioning and control of the experiment package. A roll positioning mechanism that can be driven and then locked to any position over the required roll offset range is provided.

Introduction

The Apollo Telescope Mount (ATM) will be a manned solar observatory. The broad objective of the ATM experiment is to increase our knowledge of the solar environment by observing it from a station above the major portion of the earth's atmosphere. The ATM is being developed for participation in the study of the period of maximum solar activity in 1968-69.

It is proposed to place the ATM in an earth orbit as a portion of a cluster in one of several possible configurations. The configuration under primary consideration consists of an Apollo command and service module spacecraft, an S-IVB workshop, a lunar mapping and survey system, an airlock module, a multiple docking adapter, a lunar excursion module ascent stage, and the ATM. This cluster, comprising the first four payloads of the Apollo Applications Program, will be lifted into earth orbit by four uprated Saturn I launch vehicles after which orbital assembly will take place.

The ATM pointing control subsystem (PCS) has been developed to meet the high accuracy pointing requirements established by the desired experiment conditions. These requirements must be maintained by the control system under the influence of external and internal disturbance torques, such as gravity gradient and aerodynamic disturbances and onboard astronaut motion.

Marshall Space Flight Center is responsible for directing the development of the experiments and integrating the entire system for NASA's Office of Manned Space Flight. Astrionics Laboratory is responsible for the ATM technical program. Pointing control is a portion of that program.

System Requirements

For the pointing control subsystem design requirements, roll is defined as the angular rotation about the line of sight from the experiment package to the center of the sun, and pitch and yaw are the small angular deviations of the experiment package with respect to this line of sight. The design requirements are as follows.

1. Command pointing requirements:

   a. Roll command position (ϕ_r): ±10 arc min.

   b. Pitch and yaw command position ($\phi_{p,y}$): ±2.5 arc sec.

2. Control system pointing and stability requirements about the commanded reference point:

   a. Pitch and yaw attitude: ±2.5 arc sec for 15 min of operation.

   b. Pitch and yaw rate (max. jitter rate): ±1 arc sec/s.

   c. Roll excursion: ±7.5 arc min for 15 min of operation.

   d. Roll rate (max. jitter rate): ±1 arc min/s.

   e. Maximum acquisition time: 10 min.

   f. Offset pointing:

      (1) Pitch and yaw repositioning capability from any point to any other point within a ±20 arc min square nominally centered on the solar disc.

      (2) Roll repositioning capability of ±90 degrees from the North Ecliptic Pole solar reference position.

      (3) Time not to exceed one minute, including settling time at the new attitude, for an offset maneuver in either pitch and yaw or roll.

   g. Trim resolution: manual trim to within ±2 arc sec. Where applicable, these values are considered to lie within the 1σ probability bounds and are to be achieved in the presence of nominal expected astronaut motion during intervals of experiment data gathering only.

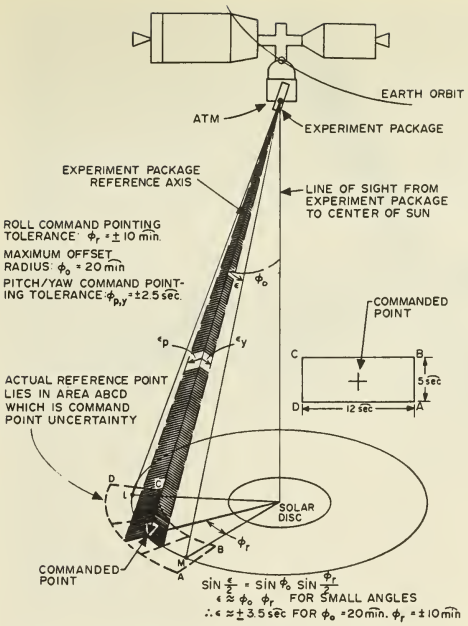


FIGURE 1. COMMAND POINTING REQUIREMENTS.

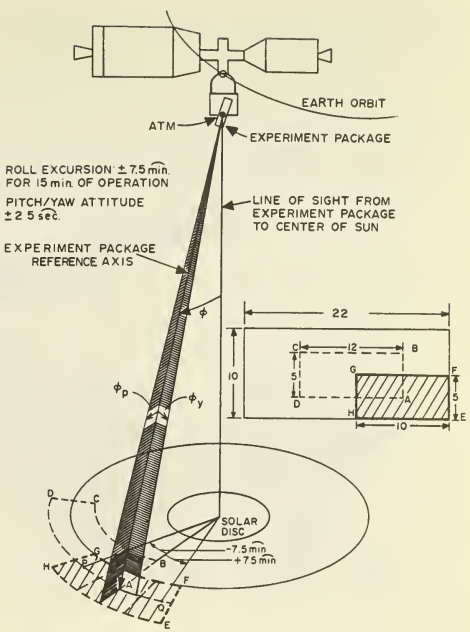


FIGURE 2. CONTROL REQUIREMENTS AROUND ACTUAL REFERENCE POINT (15-min period).

The command pointing requirements are illustrated in Figure 1 which portrays the experiment package reference axis offset from the line of sight to the center of the sun by the maximum offset angle ϕ_o of 20 arc min. At this offset angle and because of the ±10 arc min uncertainty of the roll angle ϕ_r, the commanded point in the plane of the solar disc will lie anywhere along arc length LM which for small angles may be approximated by $\epsilon = \pm 3.5$ arc sec. If in addition the ±2.5 arc sec uncertainty of pitch and yaw are included, the actual commanded reference point will lie anywhere within region ABCD, which is the command point uncertainty (Fig. 1). Since all angles are small, area ABCD may be considered to be a rectangle with a width of 5 arc sec (illustrated by ϵ_p) and a length of 12 arc sec (illustrated by ϵ_y) at an offset angle ϕ_o = 20 arc min.

The control system pointing requirements around any actual reference point are illustrated in Figure 2, where point A is considered as the actual commanded reference point. Arc length PQ in the plane of the solar disc is generated by the ±7.5 arc min roll excursion during a 15-min time interval with a 20 arc min offset from the center of the sun. It may be approximated by an angle ϕ_y of ±2.5 arc sec. The inclusion of pitch and yaw uncertainty of ±2.5 arc sec generates the area EFGH, which is the locus of points for the control of the experiment package reference axis around the commanded reference point A. Since all angles are small, area EFGH may be approximated as a rectangle with a width of 5 arc sec (ϕ_p) and a length of 10 arc sec (ϕ_y), again assuming a 20 arc min offset angle. Since the center of area EFGH can be any point within area ABCD, the command pointing uncertainty may be combined with that for the control around an actual reference point. This results in an overall experiment package pointing uncertainty which may be approximated by a rectangle with a width of 10 arc sec and a length of 22 arc sec as illustrated in Figure 2.

ATM Configurations

All configurations considered for the ATM system have been developed to make maximum possible use of available Saturn/Apollo equipment. The ATM mission module is made up of a lunar module (LM) ascent stage with an ATM rack in place of the descent stage. The optical equipment for performing the solar astronomy experiments while in earth orbit is housed in an experiment package which is gimbaled within the rack. This system consists of a roll positioning mechanism (RPM) to provide roll repositioning capability and a two-axis gimbaled experiment pointing system (EPS) to provide pointing accuracy and offset pointing of the experiment package in pitch and yaw. Control of the rack is achieved by a system using three two-degree-of-freedom control moment gyros (CMG's) to supply three-axis vehicle restoring torques. This CMG control system is a redesign of a prototype CMG which has been under investigation by the Langley Research Center for some time.[1] It is presently under development by the Eclipse-Pioneer Division of the Bendix Corporation.[2]

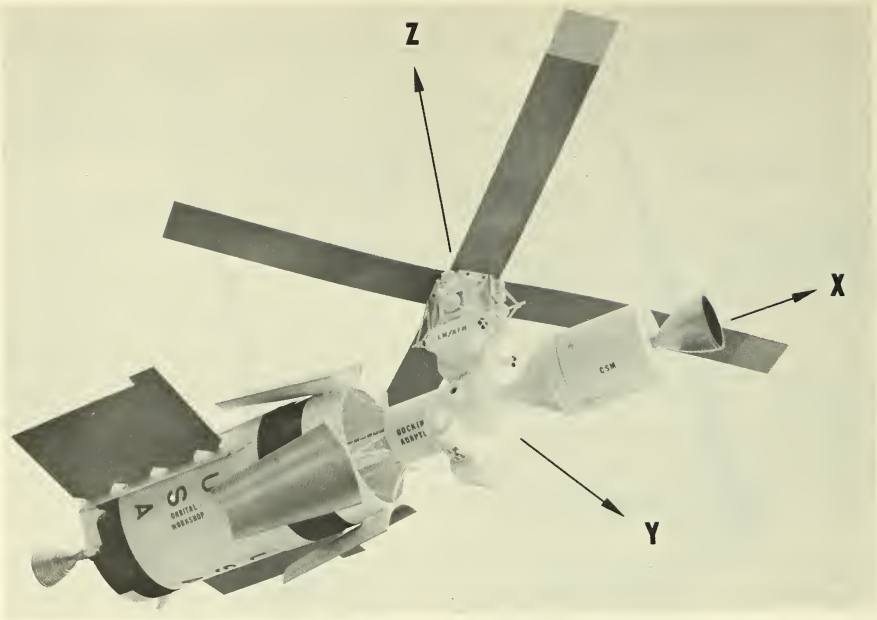


FIGURE 3. ATM CLUSTER CONFIGURATION.

The primary ATM configuration, called the cluster configuration as illustrated in Figure 3, consists of the LM/ATM docked to an aggregation of other flight modules consisting mainly of the command and service modules (CSM), the S-IVB stage of the Saturn booster, an airlock module, and the multiple docking adapter module (MDA). As illustrated in Figures 4A and 4B, the LM/ATM must also be designed to operate docked to the CSM or as a free flying LM/ATM. For all configurations, the astronaut control console is located within the LM.

FIGURE 4A. ATM COMMAND AND SERVICE MODULES FREE FLYING CONFIGURATION.

FIGURE 4B. ATM FREE FLYING CONFIGURATION.

A fourth configuration to be considered is the tethered configuration illustrated in Figure 5. For this configuration, the LM/ATM is connected to the S-IVB/MDA/CSM combination with either a rigid or a soft tether. The S-IVB/MDA/CSM is passively stabilized, primarily using the gravity gradient field, and the LM/ATM is actively controlled to point toward the sun using the CMG control system. Less than optimum performance is acceptable for the tethered configuration since no compromises are to be made in the LM/ATM system design to enable operation in this configuration.

FIGURE 5. ATM TETHERED CONFIGURATION.

As presently conceived, the primary mission plan is to use four uprated Saturn I vehicles to place the elements of the cluster configuration into an orbital altitude of approximately 485 kilometers. Launched several days apart, the first two launch vehicles of the Apollo Applications Program will place a manned Apollo spacecraft containing a mapping and survey system into a low earth orbit and an S-IVB orbital workshop into a 28.5 degree inclined orbit with a guaranteed lifetime in excess of one year (these missions are separate from the ATM mission and are not discussed). The first Apollo Applications Program spacecraft will transfer to the higher orbit of the S-IVB workshop, rendezvous, and attach the mapping and survey system to the docking adapter of the workshop. The workshop will be outfitted for living and laboratory quarters with an initial mission lifetime of up to 28 days.

Three to six months later, a third vehicle will launch a CSM plus resupply constituents to an orbit slightly below that of the workshop. A fourth vehicle will boost the LM/ATM to an orbit allowing rendezvous and docking by the CSM. The CSM will then boost the LM/ATM to the S-IVB workshop after which the CSM will undock from the LM/ATM and dock longitudinally into the multiple docking adapter attached to the workshop. Using its existing reaction control system propulsion, the LM/ATM will dock into the side of the multiple docking adapter, thereby completing the cluster configuration as illustrated in Figure 3. After a 56-day mission, the CSM will return the astronauts to earth. The LM/ATM/S-IVB workshop will then be placed in an orbital storage phase to permit a revisit and reactivation up to a year later for another 56-day ATM mission.

The free flying LM/ATM and tethered mode options will be considered feasible only during the first 56 days of operation. The duration of the free flying LM/ATM mission will be limited to a maximum of three days.

Control Philosophy

The ATM pointing control subsystem that has evolved to date has been influenced by a number of factors. The prime requirement is to meet the high accuracy system pointing specifications within the constraint of the various required configurations and the corresponding external and internal disturbance torques. The significant disturbance torques of interest are those caused by earth orbital environmental influences (gravity gradient and aerodynamic disturbances) as well as internal movements of the astronauts on board. Because of these earth orbital environmental influences, the vehicle attitude for the cluster configuration (which has the highest values for moments-of-inertia) must be held to a fixed position relative to the orbital plane. To significantly reduce constant gravity gradient bias torques, the vehicle's principal axis of minimum moment-of-inertia must be constrained to lie as close as possible to the orbital plane while the ATM experiment package points toward the sun. Since this constrains the vehicle attitude about the line of sight to the sun, the roll repositioning requirement is obtained by a roll positioning mechanism that can be driven ±95 degrees relative to the rack and then locked to any position within this offset range. To meet the pitch and yaw pointing accuracies, a two-axis gimbaled EPS with a maximum range of ±3 degrees is required. The primary requirement for the EPS is to provide experiment package isolation from the relatively large vehicle perturbations that can result because of nominal astronaut motion effects. Thus, the experiment package is gimbaled with respect to the rack by the EPS to accomplish the primary pointing task while the CMG control system compensates for astronaut disturbances and all other disturbance effects.

The CMG control system of the rack was chosen primarily because of performance benefits with respect to both dynamic response and compensation of cyclic external disturbance torques caused by gravity gradient and aerodynamic effects. Most passive control schemes (gravity gradient, for example) would not have the required accuracy and could not develop sufficient torque to meet the dynamic performance requirements. During data gathering intervals when experiment optics are exposed, use of CMG's prevents optics contamination that would result from reaction control thruster exhaust, particularly in the LM reaction control system because four of the engines exhaust directly toward the experiment package. In addition, for some of the configurations, the relatively high minimum impulse drift rates (at least 0.02 deg/s for the free flying LM/ATM configuration, for example) do not meet the required rate accuracy specified. Thus, the existing reaction control systems of the LM

and CSM (each consisting of a complement of 445 N thrusters) are available as a coarse control system. These systems will provide CMG momentum desaturation capability as well as attitude control for all other mission phases not requiring the ATM pointing control subsystem, such as rendezvous, docking, undocking, and station keeping (if required).

Pointing Control Subsystem Description

The pointing control subsystem of the ATM consists of the CMG control system, the EPS, the RPM, and associated sensors and electronics.

The CMG system is a momentum exchange control system. The momentum exchange devices are three orthogonally-mounted double-gimbaled CMG's, as illustrated in Figure 6; each has a stored momentum capability of 27 N-m-s. The gimbal rates portrayed in this

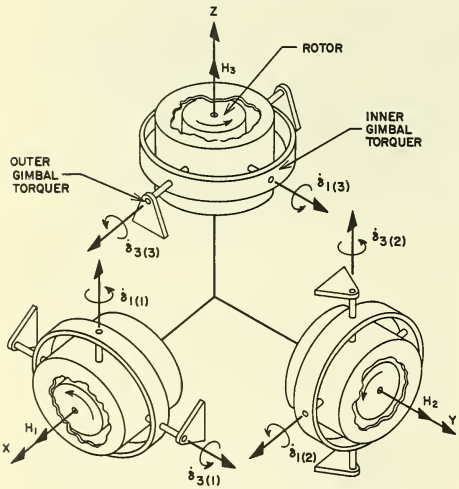


FIGURE 6. CONFIGURATION OF THE CONTROL MOMENT GYROS

figure are defined in Reference 2. While the primary approach (considering the cluster configuration only) is described, alternate approaches are under study. During the daylight portion of the orbit, cluster pitch and yaw attitude angles (rotations about the vehicle X and Y axes (Fig. 3)) are sensed by an acquisition sun sensor. During the night portion, these angles are obtained by integrating body-fixed rate gyro output signals. Pitch and yaw rate damping for the CMG control system is obtained by differentiating the acquisition sun sensor outputs during the day and from the body-fixed rate gyros at night. Cluster roll attitude angle (rotation about the vehicle Z axis (Fig. 3)) is generated by integrating the output signal of a body-fixed roll rate gyro. The CMG control system maintains the cluster in a solar-fixed attitude with its Z axis pointed toward the sun. This system can compensate for disturbance torques indefinitely only so long as the torques are purely cyclic. However, the CMG control system requires periodic momentum desaturation since the disturbance torques being compensated have a bias level. The momentum desaturation schemes utilizing the existing RCS and external disturbance torques are being considered.

The EPS consists of pitch and yaw "flex-pivot" gimbals;* each is actuated by two redundant torque motors and driven by the error signal developed from the fine sun sensor and rate gyros.[3] The EPS provides isolation of the experiment package from the unexpected man-motion disturbance torques. The CMG control system is not capable of meeting the pointing requirements in the presence of these torques; therefore, the EPS is added. The RPM provides roll attitude positioning of the experiment package through a roll offset drive motor attached to the supporting ring of the experiment package. The astronaut can "crank-around" the experiment package through ±95 degrees. This roll freedom is required to align optical "slits" of certain experiments with the limb of the sun while maintaining the desired orientation of the principal axis of minimum inertia with respect to the orbital plane.

The pointing control subsystem makes use of various sensors and sensor signal processing in the calculation of the vehicle "attitude errors." The sensors and locations are as follows.

Sensor	Location	Output
1. Acquisition sun sensor	Rack	Cluster pitch and yaw attitude
2. Fine sun sensor	Experiment package	Experiment package pitch and yaw attitude with respect to solar disc
3. Canopus tracker	Rack	Cluster roll attitude with respect to the sun's line of sight
4. Integrators on output of rate gyros	Rack	Cluster pitch and yaw attitude
5. Rate gyros	1 ea on rack	Cluster roll rate
6. Rate gyros	2 ea on experiment package	Experiment package pitch and yaw rates when EPS is active. Cluster pitch and yaw rates when EPS is caged

The acquisition sun sensor is mounted on the rack and provides pitch and yaw attitude errors of the cluster

*A flex-pivot gimbal bearing is made of a pair of flat cross-leaf springs, is welded to and supported by rotating sleeves, has no backlash, and provides limited angular travel. A preliminary analysis of an EPS concept using flex-pivot gimbals has been performed by the Apollo Applications Program study group of the Lockheed Missiles and Space Company at Sunnyvale, Calif.

with respect to the sun line. The sensor provides a total field-of-view of ±9 degrees, although its linear range is only ±5 degrees (each axis). A sun presence signal is provided from the sensor when the sun is within the ±9 degree range. The specified positioning accuracy of the sensor is ±2 arc min.

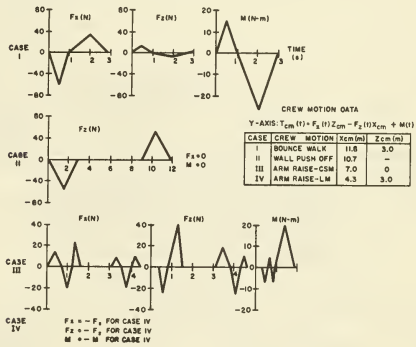
The fine sun sensor, mounted on the experiment package, provides the error signals used to drive the EPS gimbals. The 2σ accuracy of the sensor is approximately ±1.5 arc sec. The experiment package offset capability for each axis is developed by the movement of an optical wedge between the sensor and the line of sight to the sun. The position of the wedge is commanded by the astronaut's control stick. In this manner the experiment package axis is offset to any point on the solar disc within the offset capability of ±24 arc min. The solar disc is approximately 32 arc min from limb to limb so that the experiment package may be offset pointed beyond the sun's limb. Each optical wedge is provided with a digital readout so that the astronaut can determine precisely the offset attitude.

The Canopus star tracker, mounted on the rack, is used to meet the ±10 arc min accuracy requirement of the experiment package roll reference computation with respect to Ecliptic North. In addition it compensates for long term roll reference gyro drift and can provide an attitude reference for calculating the required roll command to keep the cluster's principal axis of minimum inertia close to the orbital plane. The tracker is mechanically gimbaled in two axes through ±20 degrees (inner pivot) and ±80 degrees (outer pivot). Since Canopus is off the Ecliptic South Pole by 14.5 degrees, an inner gimbal freedom of ±20 degrees was selected to permit Canopus acquisition and lock at any time during the year when the line of sight is not occulted by the earth. The outer gimbal freedom of ±80 degrees was dictated by the constraint of maintaining the vehicle's principal axis of minimum inertia approximately in the orbital plane, by the orbital plane parameters, by the celestial geometry, and by cluster orbital assembly alignment tolerances. The tracker has ±1/2 degree field-of-view in the acquisition mode and may either acquire the star automatically or be driven in a manual search mode by the astronaut. Once star acquisition has occurred, the field-of-view is reduced to ±10 arc min.

During occultations of the sun by the earth, the pointing control subsystem reverts to rate gyro control, and the attitude signals to the CMG's are determined by integrating the rate gyro output signals. This scheme is also used for roll axis attitude control most of the time because star tracker update is available only when Canopus lock-on is obtained. The integrators, located in the ATM control computer, are high gain operational amplifiers with temperature-stable feedback capacitors. The drift of these integrators is not expected to exceed 50 mV/hr, and the drift of the overall vehicle attitude reference is expected to be less than 0.5 deg/hr.

Two gyros are located on the experiment package to detect pitch and yaw rates. An additional rate gyro is located on the rack to detect roll rate. These gyros may be operated in either a coarse or a fine mode, where the respective maximum output rates are ±1 deg/s and ±0.045 deg/s. Drift characteristics are 0.033 deg/hr (1σ) for short term and 0.25 deg/hr (1σ) for long term. The requirement that dictated the choice of this type of gyro is the desired roll axis stabilization of 7.5 arc min for 15 min of time. For hardware symmetry (and thus interchangeability) the same type of rate gyro will be used for all three axes.

In the preliminary design of the ATM, the experiment package was assumed to be hard mounted to the "rack" structure. Precision pointing of the experiment package was to be accomplished by the CMG control system. It was soon realized, however, that man-motion disturbances would tax the capability of the CMG control systems to maintain the pointing stability required by the experiments. Extensive simulation studies of man-motion effects have been performed by the Apollo Applications Program study group of the Martin Company in Denver, Colo.[4,5] Typical force and moment waveforms resulting from these studies are shown in Figure 7, and typical results are shown in Figure 8.



CASE	CREW MOTION	Xcm (m)	Zcm (m)
I	BOUNCE WALK	11.8	3.0
II	WALL PUSH OFF	10.7	–
III	ARM RAISE-CSM	7.0	0
IV	ARM RAISE-LM	4.3	3.0

FIGURE 7. TYPICAL CREW MOTION DISTURBANCES.

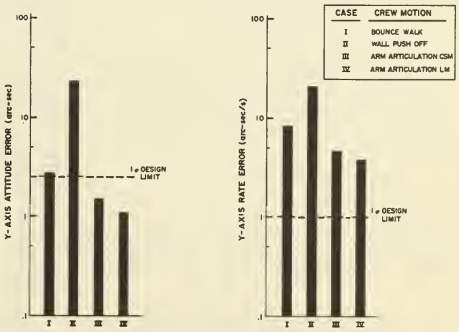


FIGURE 8. TYPICAL POINTING ERRORS FOR ASSUMED CREW MOTIONS.

Pointing Control Subsystem Operation

Modes of operation of the ATM control system are experiment pointing, monitor and acquisition, inertial hold and maneuver, and momentum dump. An initial arbitrary attitude in space following injection into orbit is assumed.

Table I indicates where the EPS and CMG control systems obtain their pitch, yaw, and roll information during the various operational modes. The first operation to be performed by the ATM system, after injection into orbit and configuration assembly have taken place, is sun acquisition. For this discussion a cluster configuration is assumed (Fig. 3). The cluster is maneuvered by the CSM reaction control system, under astronaut and CSM sensor control, until the ATM solar panels are perpendicular to the solar vector and the principal axis of minimum inertia of the vehicle lies approximately in the orbital plane. Solar vector orientation is necessary during the spin-up phase to meet the CMG cluster and associated electronics power requirement of 700 watts peak for a duration of eight hours. During spin-up, the CMG cluster is referenced to a zero angular momentum position (Fig. 9) to significantly reduce the introduction of CMG disturbance torques into system dynamics. The EPS is caged, and CMG control system rate integrators are reset. Vehicle attitude control is maintained by the CSM reaction control system.

The second operational phase begins after CMG spin-up has been completed. It consists of switching cluster control from the CSM reaction control system to the rack-mounted CMG control system. This phase will nominally take place during the daylight phase of the orbit. The astronaut will check the sun presence indicator to insure that the sun is within field-of-view (±9 degrees) of the acquisition sun sensor. This sensor provides error signals (Fig. 10) which, when nulled by the CMG control loop, position the ATM cluster Z axis to within approximately ±2 arc min from the line of sight to the center of the solar disc. With the sun presence sensor activated,

TABLE I. PRELIMINARY ATM MODE DEFINITION

MODES	VERNIER SYSTEM			FINE SYSTEM*		
	Pitch and Yaw Attitude	Roll Attitude	Pitch and Yaw Rate	Pitch and Yaw Attitude	Roll Attitude	Pitch and Yaw Rate
Experiment Pointing Mode[1]	Fine sun sensors on experiment package, error signal from optical wedges set by control stick	Roll positioning mechanism set by control stick unless override switch is on; the roll positioning mechanism is locked	Rate gyros	Acquisition sun sensor	Integrate rate gyro unless override is on, then integrate control stick and rate gyro.	Lead network on output of acquisition sun sensor
Monitor and Acquisition Mode Day[2]	EPS gimbals caged at zero; wedges zeroed	Locked at last position	Rate gyros active, resolved to rack coordinates	Acquisition sun sensor	Integrate rate gyro	Resolved experiment package rate gyros.
Night[3]	EPS gimbals caged at zero; wedges zeroed	Locked at last position	Rate gyros active, resolved to rack coordinates	Resolved integrated experiment package rate gyros	Integrate rate gyro	Resolved experiment package rate gyros
Inertial Hold and Maneuver Mode	EPS gimbals caged at zero; wedges zeroed	Locked at last position	Rate gyros active, resolved to rack coordinates	Resolved integrated experiment package rate gyros only or rate gyros and control stick	Integrate rate gyro only or rate gyro and control stick	Resolved experiment package rate gyros
Momentum Dump Mode[4]	EPS gimbals caged at zero; wedges zeroed	Locked at last position	Rate gyros active, resolved to rack coordinates	Resolved integrated experiment package rate gyros	Integrate rate gyro	Resolved experiment package rate gyros

*Roll rate is obtained from the rate gyro.

[1] Automatically switch to night monitor and acquisition mode at clock time.

[2] Automatically switch to night side hold at clock time.

[3] Automatically switch to day at clock time.

[4] Attitude error signals to CMG control system open.

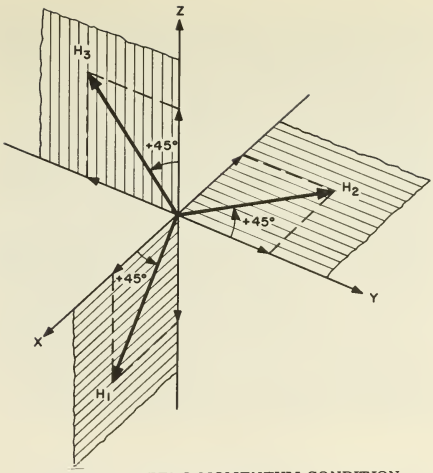


FIGURE 9. ZERO MOMENTUM CONDITION.

the astronaut switches to "monitor and acquisition mode" control. As previously indicated, pitch and yaw attitude error information is supplied by the acquisition sun sensor. Pitch and yaw attitude rate information is obtained by differentiating the attitude error signals. Roll rate is obtained from the roll rate gyro, and roll attitude is determined by integrating the rate gyro output.

The third operational phase is concerned with locking the star tracker on Canopus. While the ATM cluster is controlled by the pointing control subsystem in the monitor and acquisition mode (probably the most likely operational mode when acquiring Canopus for the first time), the star tracker is placed in its drive mode (SI closed, Fig. 10). This permits the tracker to be positioned by applying torque about its inner or outer gimbal axes. Since the inner gimbal position is a function of the time of year, it can be set to the desired position. The outer gimbal is then driven in a manual search mode until a discrete signal indicates that Canopus is within the tracker field-of-view of ±1/2 degree. The tracker is then switched (SI opened, Fig. 10) to an automatic tracking mode, and the control stick output is

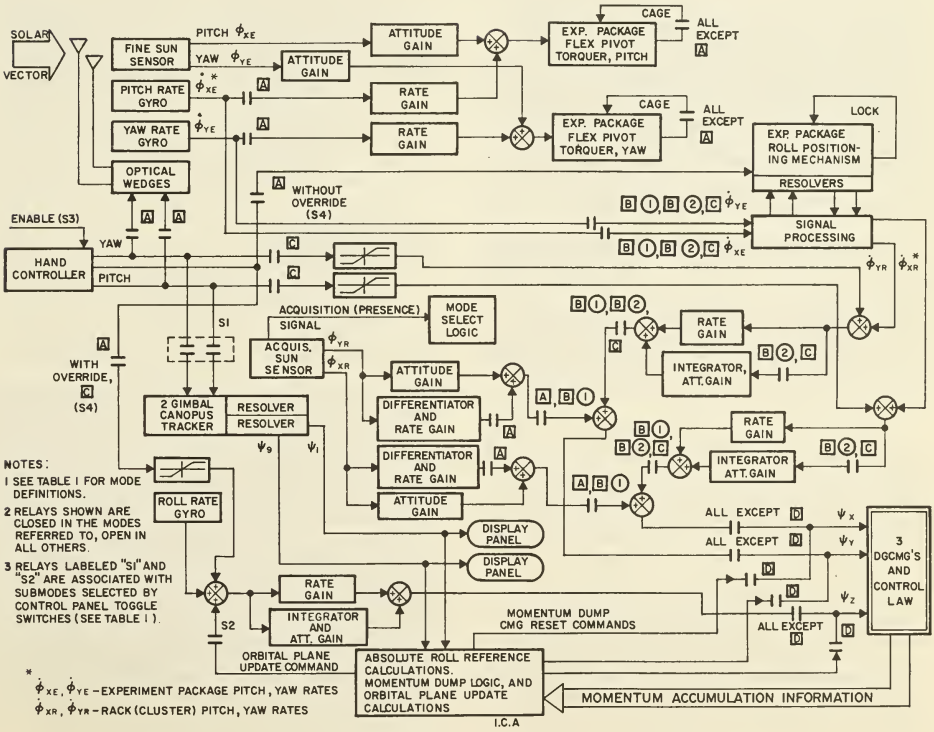


FIGURE 10. BLOCK DIAGRAM OF THE ATM CONTROL SYSTEM MODE DEFINITION.

uncoupled from the tracker. In the event Canopus is occulted or lost, the tracker gimbals are locked until Canopus comes out of occultation or the tracker is switched to the track mode. With the star tracker locked on Canopus and the acquisition sun sensor (replaced by the gyro reference during night intervals) locked on the center of the solar disc, the roll reference rate gyro can be updated and an orbital plane update command can be issued during non-data gathering intervals at the astronaut's discretion (close switch S2, Fig. 10). Thus, the vehicle's principal axis of minimum inertia lies approximately in the orbital plane.

These three operational phases nominally occur only during the initial phase of the ATM mission before solar experiment data gathering has begun. System operation during phases other than these are best described by discussing each of the basic modes of control.

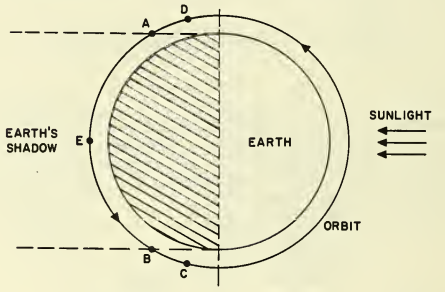
The experiment pointing mode is the operational mode used during periods of data gathering. The CMG control system maintains the cluster attitude solar oriented with the vehicle Z axis pointed at the sun. The EPS is driven by the fine sun sensor error signal. Offset command capability of the EPS is provided by pitch and yaw optical wedges which are rate commanded by the astronaut's control stick. In this mode the roll channel of the control stick normally commands the roll positioning mechanism of the experiment package. The astronaut has the capability of overriding this condition so that the roll channel of the control stick commands the roll axis of the CMG control system. This might be desirable in case a roll attitude of the experiment package was required beyond the ±95-degree RPM offset range. The astronaut could then obtain the desired attitude by a small maneuver of the entire cluster.

It is expected that periods will exist during daylight operation for which the experiments need not be in operation because of solar inactivity or the astronaut's performing other functions. The system will then be placed in the monitor and acquisition mode. Both the EPS and the RPM will be caged, and the CMG control system will maintain the cluster attitude in inertial hold. This mode can exist during day and night cycles, as indicated in Table I.

In the inertial hold and maneuver mode, the astronaut has the capability of maneuvering the entire cluster, using the CMG's. In this mode the EPS and RPM are caged. The CMG control system maintains an inertial hold unless the astronaut commands an attitude maneuver. If the astronaut places the cluster in a given attitude, the CMG control system will maintain that attitude.

In the momentum dump mode, the EPS and RPM are again caged; the CMG control system operation is the same as that of the night monitor and acquisition mode (Table I). Although attitude error signals are available from the integrators in all three axes, they are not sent to the CMG control system. This causes any attitude perturbation existing after the momentum desaturation period to be taken out when the CMG loop is again closed.

An important aspect of the pointing control subsystem operation is mode selection and automatic mode switching. Mode selection is accomplished by manually positioning the mode selection switch to the desired mode of operation. In certain operational modes, automatic mode switching also occurs. Automatic switching from the experiment pointing mode to the monitor and acquisition mode (night interval) is initiated by a discrete signal which occurs approximately at orbital sunset. Automatic switching between the monitor and acquisition mode (night interval) and the monitor and acquisition mode (day interval) then occurs until the mode selection switch is positioned to a new mode. The discrete signal used to initiate day or night mode switching is supplied by an interval timer. The signal is issued several minutes before loss of the sun when going into the night cycle and again several minutes after sun acquisition when coming into the daylight cycle to prevent atmospheric solar refraction effects from placing transients on the ATM pointing control subsystem. Figure 11 indicates the desired discrete issue points "C" and "D."



A = LOSS OF SUN
B = SUN PRESENCE
C = DESIRED POINT IN ORBIT TO SWITCH TO DAY-SIDE MONITOR CONFIGURATION
D = DESIRED POINT IN ORBIT TO SWITCH TO NIGHT-SIDE HOLD CONFIGURATION
E = "MIDNIGHT"

FIGURE 11. INTERVAL TIMER REQUIREMENTS.

Passive Gravity Gradient Stabilization Aspects of ATM

The present ATM mission identifies two possible applications for passive gravity stabilization. The first is the rigid tether or boom arrangement shown in Figure 5, which is an alternate configuration for the active part of the mission. The LM/ATM is attached to an extendable member by a ball joint and deployed approximately 30 meters from the multiple docking adapter. The composite vehicle is gravity gradient stabilized with the boom axis nominally along the orbital radius vector and the engine of the CSM is in the direction of the flight path. The LM/ATM continually points at the sun and hence rotates with respect to the rest of the vehicle. In general the boom axis will not align with the local vertical since there is some non-zero aerodynamic torque in this position and since the boom axis is probably not a principal axis. Therefore a bias angle of a few degrees off the local vertical in the plane of the orbit is expected. This angle perhaps can be minimized by adjusting the boom

length. A passive damping mechanism presently under consideration is the General Electric magnetically-anchored, viscous damper.[6] This particular damper was considered because of its compactness and flexibility in being integrated into the spacecraft. The damper is composed of two concentric spheres separated by a viscous fluid. The inner sphere contains a large permanent magnet which aligns itself to earth's magnetic field, and the outer sphere is attached to the spacecraft. Relative motion between the two bodies causes energy to be dissipated. This type of damper does not require long, articulated appendages which cause mounting problems and could possibly interfere with solar arrays or astronaut extravehicular activity. The disturbance torque arising from a non-fixed reference, such as the earth's magnetic field, is usually undesirable but may not be significant in this application. Since the natural frequencies in two of the axes for the linearized equations very nearly correspond to orbital and twice orbital rate, the steady-state amplitudes may be relatively large. This problem presently is under investigation.

The second application for passive gravity stabilization is during ATM storage. During this period, the vehicle has the form shown in Figure 12. While the vehicle is in the storage mode, there is a requirement to maintain certain temperature ranges in the spacecraft and to keep telemetry channels active. To meet this requirement, the solar panels must be oriented to guarantee a given amount of power; therefore, the gravity-stabilized system was proposed, since this system orients the panels so that they are illuminated for one quarter of each orbit. The stable attitude for this vehicle is that in which the longitudinal axis of the S-IVB is nominally in the direction of local vertical, and the vehicle Z axis is pointing in the opposite direction of the velocity vector. Since the geometric axes are not, in general, principal axes and the aerodynamic torques are significant, the vehicle X axis will actually be biased off the local vertical in the orbital plane. There will be very little gravity gradient restoring torque about the axis of minimum moment-of-inertia because the other two principal moments are almost equal. However, the LM/ATM with its solar arrays acts like an aerodynamic rudder and will tend to keep the plane containing the vehicle X and Z axes in the orbital plane. The passive dampers for this configuration are those previously described. It is expected that final steady-state amplitudes can be kept less than 10 degrees about the reference in all axes.

FIGURE 12. ATM STORAGE MODE CONFIGURATION.

Conclusions

Based on the mission and high accuracy experiment pointing requirements, a preliminary design for an ATM pointing control subsystem has been developed to meet the experiment objectives for any of the various vehicle configurations required. A significant portion of the pointing control subsystem is capable of manual operation. Thus, astronaut operation is available to perform various functions such as those which require judgment to select scientific targets and to point experiments toward these targets. To keep the number of tasks required by the astronaut within reason, many functions such as attitude pointing stability are performed using closed loop automatic control. However, because of the manual control capability and television optics display system, the astronaut can manually improve some of these functions such as offset pointing if sufficient display accuracy and resolution are available.

In addition to the experiment pointing requirements, the cluster vehicle configuration as well as disturbance torques resulting from man-motion effects were dominant considerations affecting the pointing control subsystem design that has evolved to date. The high values for moment-of-inertia of the cluster placed a constraint on the vehicle attitude about the line of sight to the sun primarily because of gravity gradient torques. This led to the design of the RPM to meet the roll repositioning requirement of ±90 degrees. Because of vehicle perturbations resulting from nominal astronaut motion effects, the requirement for a two-axis gimbaled EPS was established to provide experiment package isolation from these effects.

Overall attitude control of any of the available vehicle configurations is then obtained from a CMG control system contained within the LM/rack combination. A system consisting of three two-degree-of-freedom control moment gyros supplies vehicle restoring torque control. This system was selected primarily because of performance benefits with respect to dynamic response, efficient compensation of cyclic disturbance system torques, and avoidance of possible optics contamination from the reaction control exhaust. The investigations by the Langley Research Center and the development effort by the Bendix Corporation aided in the selection of this system.

Four primary modes of operation of the ATM control system have been developed for the various operational phases of the mission. Mode selection is primarily accomplished by manually positioning mode selection switches located on the display and control console in the LM ascent stage. As described, however, mode switching is effected automatically for some of the operational modes.

It must be recognized that much of the data and characteristics used in the development of the preliminary

pointing control subsystem will continually change as the experiment requirements as well as vehicle configurations become finalized. As a result, all analyses and simulations will be continuously updated and alternate approaches will be developed for incorporation into the design using the hardware presently under development.

References

1. Kurzhals, Peter R., and Grantham, Carolyn; A System for Inertial Experiment Pointing and Attitude Control, NASA TR R-247, NASA Langley Research Center, Langley Station, Hampton Virginia, 1967

2. Morine, Louis A., and O'Conner, Bernard J.; The Description of the CMG and Its Application to Space Vehicle Control, The Bendix Corporation, Navigation and Control Division, Teterboro, New Jersey, 1967.

3. Conceptual Design and Analysis of a Control System for Apollo Telescope Mount, Lockheed Missiles and Space Company, Sunnyvale, California, 1967

4. Apollo Applications Program (AAP) Payload Integration Technical Study and Analysis Report, Single Axis Hard Mounted ATM Control System Study, Martin Marietta Corporation, Denver, Colorado, 1967

5. Apollo Applications Program (AAP) Payload Integration Technical Study and Analysis Report, Single Axis Gimballed ATM Control System Study, Martin Marietta Corporation, Denver, Colorado, 1967

6. Katucki, R. J., and Moyer, R. G.; Systems Analysis and Design of a Class of Gravity-Gradient Satellites Utilizing Viscous Coupling between Earth's Magnetic and Gravity Fields, Spacecraft Department, General Electric Company, published in notes of Symposium on Passive Gravity-Gradient Stabilization, Ames Research Center, Moffett Field, California, 1965

No. 67-535

PERFORMANCE CHARACTERISTICS OF A "DAMPED" MASS EXPULSION ATTITUDE CONTROL SYSTEM

by

W. C. ENGLEHART
Aerospace Corporation
El Segundo, California

AIAA Paper
No. 67-535

AIAA Guidance, Control and Flight Dynamics Conference
HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.01, 7.03, 7.04

-- NOTES --

PERFORMANCE CHARACTERISTICS OF A "DAMPED" MASS EXPULSION ATTITUDE CONTROL SYSTEM

W. C. Englehart
Member of the Technical Staff
Aerospace Corporation
El Segundo, California

Abstract

A new mass expulsion approach to space vehicle attitude control is introduced which shows promise of reducing propellant consumption and pulsing frequency plus providing better acquisition performance (time and propellant)and attitude accuracy than present minimum-impulse controllers. This "damped" attitude control logic is designed to optimally adapt to existing disturbance torques in damping the vehicle's position and rate, thereby driving the vehicle near the minimum fuel expenditure trajectory during each control cycle. Simple adaptive compensation based on the immediate past history of control pulses is included to correct for biased and time-variant disturbance torques, propulsive and sensing errors, and time-variant moments of inertia. The limit-cycle performance of the damped controller is compared with a pseudo-rate and a minimum-impulse controller in the presence of propulsive errors, sensing errors, and external disturbance torques.

I. Introduction

As mission durations are increased and mission requirements become more demanding, the problem of obtaining efficient and economical attitude control system performance becomes one of paramount importance. The new attitude control approach presented herein can significantly reduce the number of pulses and the amount of attitude control propellant used by presently envisioned spacecraft. This approach entails a more sophisticated control logic than those presently used.

Much effort has been expended to date on obtaining methods of converging to the minimum-impulse limit cycle. The "damped" controller mechanization illustrates the fact that the minimum-impulse limit cycle is not the limiting design point for mass expulsion systems. The vehicle's motion can be damped below that of minimum-impulse controllers through appropriate control system logic. Specifically, the new "damped" control logic attempts to drive the vehicle to the minimum propellant expenditure trajectory during each control cycle by taking advantage of the existing disturbance torques. The minimum propellant expenditure trajectory is that trajectory which yields the smallest average propellant consumption rate, i.e., propellant used per firing divided by the time between firings. The control-jets are turned off at a place in the phase-plane such that the natural trajectory of the spacecraft due to the disturbance torques maintains the spacecraft within the required attitude limits for the maximum time. This, then, will be the minimum fuel expenditure and minimum pulsing trajectory. An adaptive compensation scheme is included in the controller to correct for time-variant torques and parameters.

In this paper the damped control logic is discussed, and typical performance results are presented. The damped control logic is presented for space vehicles which maintain a specified orientation in all three axes. The results were obtained from two digital computer simulations which included the effects of various types of disturbance torques, propulsion errors, and sensor errors. The first of these was a single-axis kinematic simulation with constant disturbance torque coefficients. Using this simulation, performance of the damped controller was compared to that of a pseudo-rate controller, which is one type of minimum-impulse controller. The second simulation was a three-axis dynamic simulation which included time-variant disturbance torque coefficients and vehicle dynamics. Using this simulation, the damped controller performance was compared to the performance of a simple minimum-impulse controller.

II. Discussion

Damped Controller Philosophy

The philosophy of the damped controller is to drive the vehicle's rate and position as close to the minimum propellant expenditure trajectory during each control cycle as the sensing and propulsive systems will allow. This trajectory and the performance of the damped controller are dependent upon the existing disturbance torques. The operation and performance of the damped controller will, therefore, be discussed according to the nature of the disturbance torques. Aerodynamic and gravity-gradient disturbance torques were considered in the simulations. In the three-axis dynamic simulation, these torques are cross-coupled and time-variant. However, for small attitude deviations, the disturbance torques can be described approximately by $M = K_o\theta + K_1$, which represents four types of disturbance torque: zero (K_o and K_1 equal zero); constant bias (K_o is zero); destabilizing, with or without a bias (K_o is positive); and stabilizing, with or without a bias (K_o is negative). In general, K_o and K_1 are functions of time. When a bias is present in the stabilizing or destabilizing cases, the effect is a shifting of the zero-torque axis. The disturbance torques are further discussed by Dahl, et al.(1)

The optimal nature of the damped controller can be examined according to the nature of the disturbance torques. Phase-plane trajectories for the four basic types of disturbance torque are shown in Fig. 1 without zero-torque-axis offsets. The initial pulse in each case is fired at the deadband crossing signal, point 1, and the pulse is terminated at a rate cutoff setting, $\dot{\theta}_{set}$, leaving the vehicle near point 2. The controller action, after the initial pulse, is examined below for each case shown in Fig. 1.

67-535

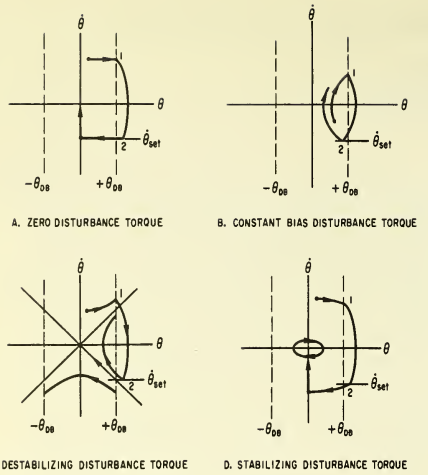


Figure 1. Phase-Plane Trajectories for Disturbance Torques Without Zero-Torque-Axis Offsets

A. Zero Disturbance Torques

A pulse is fired at zero attitude (Fig. 1A), which nulls the existing rate. Ideally, the vehicle will be driven to the equilibrium point, i.e., the origin of the phase-plane. Since the control system is not perfect, errors will exist which cause the vehicle state to be driven to some neighborhood of the origin. Thus, the vehicle will have a small residual rate and will coast slowly to the deadband. With the proper system design (rate sensor errors and rate cutoff settings), the average time spent within the deadband will be greater than the corresponding time of a minimum-impulse controller. Consequently, propellant consumption and pulsing frequency are reduced. The adaptive compensation causes the rate cutoff settings to converge to values which allow the trajectory, after a pulse firing at zero attitude, to be determined solely by system errors. Control logic which attempts to leave the vehicle with a small residual rate of the opposite sign at the deadband crossing was also investigated, but the propellant consumption was greater than that of the damped control logic due to rate sensor errors.

B. Constant Bias Disturbance Torques

The phase-plane trajectories are parabolas for constant bias torques (Fig. 1B), and the bias torques are counteracted directly by the mass expulsion system with all control mechanizations. However, the adaptive logic within the damped controller increases the pulse-width for this case since the absolute value of the rate cutoff settings is increased when the same deadband is continually crossed. Therefore, the performance is improved with the damped controller due to the reduced number of pulses and to the increase in I_{sp} that is generally obtained with longer pulse-widths.

C. Destabilizing Disturbance Torques

If the disturbance torque is destabilizing (Fig. 1C), the phase-plane trajectory will be a hyperbola. The asymptotes of the hyperbola are defined by $\dot{\theta} = \pm\lambda(\theta - \theta_{c_i})$. The adaptive compensation adjusts $\dot{\theta}_{set_i}$ so that the thrusters stop firing when the asymptote of the hyperbola family is reached. If the same deadband is continually crossed, $|\dot{\theta}_{set_i}|$ is raised. If the deadband region is continually traversed, $|\dot{\theta}_{set_i}|$ is lowered. It can be seen from Fig. 1C that this scheme will converge to the asymptote. System errors influence the region to which the vehicle state migrates. In any event, propellant consumption and pulsing frequency decrease as the vehicle state at rate cutoff approaches the asymptote. A pulse is not fired at zero attitude since the rate at zero attitude is less than $\dot{\theta}_{set_i}$.

D. Stabilizing Disturbance Torques

If the disturbance torque is stabilizing (Fig. 1D), the phase-plane trajectory will be an ellipse. A pulse is fired at zero attitude when $|\dot{\theta}| \geq |\dot{\theta}_{set_i}|$, and this pulse theoretically nulls the existing rate, $\dot{\theta}$. The vehicle state will actually be driven to some neighborhood of the phase-plane origin which will result in a limit cycle of sufficiently small amplitude that the attitude deadbands will not be reached unless the reference attitude or the nature of the disturbance torques changes. The adaptive compensation adjusts the rate cutoff settings such that the rate is nulled at zero attitude except for system errors.

E. Offset Attitude Dependent Torques

In general, the zero-torque axis (θ_{c_i}) will not coincide with the zero-reference axis. Physically this occurs when the principal axes of inertia and/or the center of mass or center of pressure are offset from the body symmetry (reference) axes. Various phase-plane trajectories are presented in Fig. 2 for large zero-torque-axis offsets.

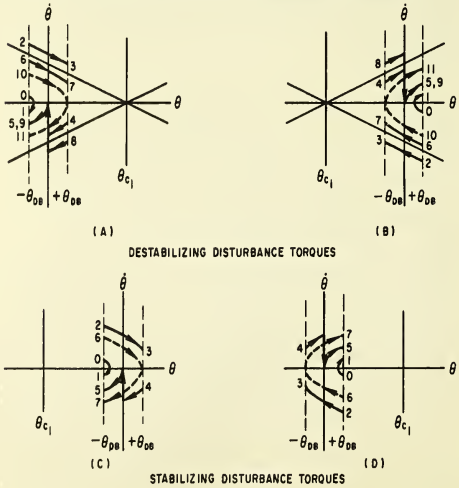


Figure 2. Phase-Plane Trajectories for Disturbance Torques With Large Zero-Torque-Axis Offsets

With only position and rate information available, the limit-cycle trajectories of curves A and C and curves B and D appear the same. Compensation (reduced disturbance torques) by moving the zero-reference axis to coincide with the zero-torque axis is impossible since the position of the zero-torque axis or the nature of the disturbance torques is, in general, not known.

Even for these cases the damped control logic can provide better performance than a minimum-impulse controller. This is accomplished as follows: when the same deadband is continually crossed, the adaptive logic increases the absolute value of the rate cutoff settings until the opposite deadband is crossed. If the region between the deadbands is continually traversed, the adaptive compensation reduces the absolute value of the rate cutoff settings. Therefore, the adaptive compensation converges the rate cutoff settings to values which allow the vehicle to cross the opposite deadband occasionally, i.e., trajectories approximating the dashed line cases of Fig. 2. The dashed trajectories are the minimum propellant expenditure trajectories in the destabilizing torque case. In the biased stabilizing torque case the dashed trajectories are not optimum; however, improved performance is realized to the extent that the increase in I_{sp} with longer firing time offsets the normal increase in impulse consumption apparent when the rate cutoff is increased.

Damped Controller Mechanization

Various mechanizations satisfying the damped controller philosophy are potentially possible, depending upon what is known about the disturbance torques and the availability of vehicle attitude and rate information. Availability of both rate and attitude data is assumed here. No specific prior information is assumed about the nature of the disturbance torques other than a rough idea of their magnitude. In the damped controller, engine firings occur at three attitudes:

1. $\theta = 0$ when $|\dot{\theta}| \geq |\dot{\theta}_{set_i}|$ (positive pulse delivered when $\dot{\theta}$ is negative and negative pulse delivered when $\dot{\theta}$ is positive)
2. $\theta \geq +\theta_{DB}$ (negative pulse delivered)
3. $\theta \leq -\theta_{DB}$ (positive pulse delivered).

Two possible mechanizations differ only in the manner of terminating engine firings. The pulse-width computation mechanization terminates the firing at the end of a period of time calculated on the basis of the rate at the deadband crossing and the existing rate cutoff setting. The rate signal cutoff mechanization terminates the firing when the rate measurement is equal to the existing rate cutoff setting.

The pulse-width computation mechanization, which minimizes bandwidth requirements for the rate sensors, is presented functionally in Fig. 3. Referring to Fig. 3, relay R1 allows a firing signal to be transmitted to the proper thruster when a deadband is crossed (switching amplifier 1). The amplitude of the signal delivered to the latching networks is proportional to the desired pulse-width since $\Delta t = (I/T)\,\Delta\dot{\theta} = K_3\,(|\dot{\theta}_{set_i}| + |\dot{\theta}|)$. Latching networks 1 and 2 latch for a time equal to the input amplitude. Relay R2 issues a firing command at zero attitude which is proportional to the rate at zero attitude. The thresholds of switching amplifier 2 are determined by the rate cutoff settings. The amplitude of the signal delivered to latching network 2 is proportional to the desired pulse-width when the rate is greater than or equal to the rate cutoff setting, and is zero when the rate is less than the rate cutoff setting. The impulse delivered at zero attitude will be approximately that required to cancel the existing rate since the engine firing

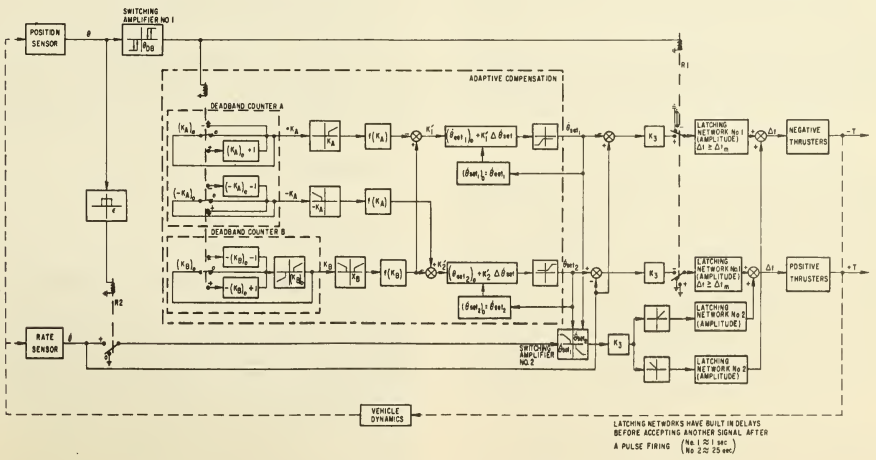


Figure 3. Damped Attitude Control Logic, Pulse-Width Computation

67-535

time is a constant, K_3, times the rate signal at zero attitude.

When the rate signal cutoff mechanization is employed, propulsion system errors and moment of inertia variations are automatically compensated for. The firing is terminated by a rate signal which indicates that the proper rate cutoff setting has been crossed. A relay is closed at zero attitude and is opened at zero rate for engine firings at zero attitude. The duration of the firing signal is the desired pulse-width.

Adaptive compensation, based on the immediate past history of deadband crossings, is included to compensate for a) propulsion and sensor errors, b) biases, uncertainties, and time-variance of disturbance torques, and c) changing moments of inertia. As seen in Fig. 3, the rate cutoff settings, $\dot{\theta}_{set_i}$, determined by the adaptive compensation influence all engine firings. The rate cutoff settings determine whether a pulse is to be fired at $\theta = 0$, which would cancel the existing rate. The desirability of not firing a pulse at $\theta = 0$ when $|\theta| < |\dot{\theta}_{set_i}|$ is based on performance characteristics. At a deadband crossing, a pulse is always delivered, its width being determined by the rate cutoff setting and the rate measurement. The adaptive compensation varies the rate cutoff setting in such a manner that the vehicle approaches the minimum propellant expenditure trajectory. Ideally, the adaptive compensation allows the damped control logic to be independent of the nature of the disturbance torques.

The adaptive compensation can be mechanized by either digital or analog elements, and includes two counters which effect nonlinear gain changes on the rate cutoff settings. The rate cutoff settings are given by

$$\dot{\theta}_{set_1} = (\dot{\theta}_{set_1})_o + K_1' \Delta\dot{\theta}_{set}$$

$$\dot{\theta}_{set_2} = (\dot{\theta}_{set_2})_o + K_2' \Delta\dot{\theta}_{set}$$

$|\dot{\theta}_{set_i}|$ is limited both on the high and low ends due to performance considerations. The lower limit for all simulations is $|\dot{\theta}_{set}|_{min} = |\dot{\theta}_{min}/2|$. Two counter networks, deadband counters A and B, provide the deadband counter signal.

Deadband counter A counts the number of consecutive lefthand or righthand deadband limit crossings, K_A. As long as the same deadband is crossed, the total number of crossings is summed. When the opposite deadband is crossed, the counter is reset and henceforth contains the consecutive sum for that deadband. $|\dot{\theta}_{set_i}|$ is increased as a function of K_A for that side of the deadband which is crossed X_A or more consecutive times. Deadband counter B counts the number of consecutive complete deadband traversals. This is accomplished by summing the signed deadband crossing (+ or - deadband) with the present accumulated count, K_B, and then changing the sign of the accumulated count. If the absolute magnitude of K_B is less than $(K_B)_o$, K_B is set to zero in order to avoid counting down. This is required since $|\dot{\theta}_{set_i}|$ is limited. If the vehicle continues to cross the same deadband, the accumulated count (K_B) will average to zero. $|\dot{\theta}_{set_i}|$ is decreased for both deadbands as a function of K_B if the deadband is traversed X_B or more consecutive times. A typical sequence of events illustrates the operation of deadband counter B. Assume the vehicle is continually traversing the deadband and $(K_B)_o$ equals zero. If the vehicle crosses the positive deadband initially, K_B becomes equal to +1. Now $(K_B)_o$ equals +1 and the threshold value $(K_B)_o$ equals ±1. The negative deadband will be crossed next and K_B becomes equal to -2. K_1' and K_2' are changed when $|K_B|$ becomes equal to or greater than X_B, and $\dot{\theta}_{set_i}$ is changed only when K_1' or K_2' changes.

The adaptive compensation presented herein does not compensate for certain parameters when zero or stabilizing disturbance torques are present. These parameters are rate sensor biases and, when the pulse-width computation mechanization is employed, the propulsion system errors and moment of inertia variations. The problem is that the rate cannot be nulled at zero attitude. If compensation for these parameters is desired, K_3 can be varied as a function of the residual rate when a pulse is delivered at zero attitude.

The initial rate cutoff settings, $(\dot{\theta}_{set_i})_o$, are determined from prior knowledge of the disturbance torques which will initially be encountered. For example, if unbiased destabilizing disturbance torques are initially expected, the rate cutoff should be set at $(\dot{\theta}_{set_i})_o = -\lambda\theta_{DB}$. The rate cutoff setting will be varied in flight by the adaptive compensation.

Three adaptive gains were investigated, the first exhibiting an extremely slow convergence while the second and third exhibit a relatively fast convergence. The f(K)'s and X's for these three gains are given below.

Gain A (Kinematic Simulation)

$f(K_A) = K_A - 1$, $f(K_B) = K_B$, and

$X_A = X_B = 4$

Gain B (Kinematic Simulation)

$f(K_A) = (K_A - 2)^2$, $f(K_B) = (K_B - 2)^2$,

and $X_A = X_B = 3$

Gain C (Dynamic Simulation)

$f(K_A) = 5(K_A - 1) - 9$, $f(K_B) = 5(K_B - 1) - 9$,

and $X_A = X_B = 3$

III. Damped Controller Performance

Typical results from two digital computer simulations are presented to illustrate the damped controller's performance. The first simulation is a single-axis kinematic simulation[2] which considers constant disturbance-torque coefficients and system errors in order to clearly illustrate the damped controller's operation. The second simulation is a three-axis dynamic vehicle simulation[3] which examines the effects of cross-coupled, time-variant disturbance torques and vehicle dynamics on the damped controller's performance.

Constant moments of inertia and propulsion parameters representative of a pulse-width modulated, hypergolic-bipropellant system were assumed in both simulations. The minimum pulse-width (Δt_m) was 0.016 sec, and the specific impulse was a function of pulse-width ($I_{sp} = 290 - 250\, e^{-20\Delta t}$, sec).

Single-Axis Kinematic Simulation Results

Propellant consumption and pulsing frequency comparisons relating the damped controller to the pseudo-rate controller[4] shown in Fig. 4 are presented for typical vehicle parameters and various constant disturbance torque coefficients (K_0 and K_1). The limit-cycle performance of the two controllers, which will exist once acquisition has been performed and the vehicle settles into a nearly repeatable limit-cycle motion, is presented.

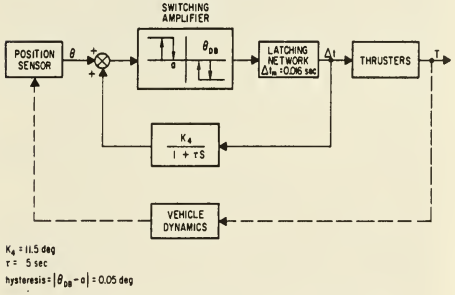


Figure 4. Pseudo-Rate Control Logic

Included in the simulation are the effects of:

1. zero-disturbance-torque axis offsets
2. uniformly distributed rate signal resolution
3. normally distributed rate signal noise
4. uniformly distributed position sensor error at zero attitude
5. reaction-jet time delay, $\tau_j = 0.01$ sec
6. normally distributed impulse error as a function of pulse-width, $\sigma = 0.0134 + 0.149\, e^{-95\Delta t}$.

Unless otherwise specified, the following nominal set of errors was used in this simulation.

RES = 1×10^{-4} deg/sec, $\sigma' = 1 \times 10^{-4}$ deg/sec, and $\theta_\epsilon = 0.1$ deg.

The error on the rate (engine cutoff) signal is obtained from a uniformly distributed random number describing the rate signal resolution and summed with a noise error picked out of a normal distribution. A thrust of 50 lbs and a T/I ratio of 0.0017 rad/sec^2 were assumed. The initial conditions were such that the vehicle had one-half the minimum-impulse rate at the initial deadband crossing, and the total mission time was approximately two days. Only propulsive errors were considered for the pseudo-rate controller. Parametric results designed to define the operation of the damped control logic for various orbital conditions are presented according to the nature of the disturbance torque.

A. Zero Disturbance Torques

The effects of rate signal noise variations on the damped controller performance are presented in Fig. 5 for zero disturbance torques.

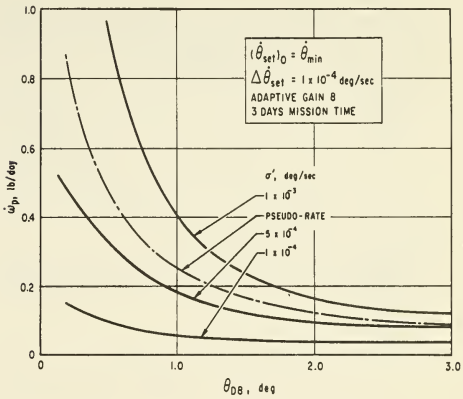


Figure 5. Effect of Rate Signal Noise Variations, Zero Disturbance Torque

Fig. 5 illustrates that the performance is directly related to rate sensor errors. Rate signal resolution variations showed the same trends as the rate signal noise variations. Varying the initial rate cutoff setting $(\dot\theta_{set})_0$ did not have a large influence on performance after three days of operation. Variations in the zero position error showed only a minor effect on performance, and these errors can actually aid the performance of the damped controller because of the combination of the random sign associated with the rate and position error. When the combined errors allow more attitude travel, the performance is improved. The effect of varying $\Delta\dot\theta_{set}$ showed no performance variation due to the extremely long coast times and the associated few engine firings.

B. Constant Bias Disturbance Torques

The performance for a large (M = +0.2 ft-lb) constant bias torque is presented in Fig. 6. The damped control logic reduced the propellant consumption by roughly one-half because the adaptive logic converged to a $\dot\theta_{set_1}$ which just allowed the opposite deadband to be crossed. By increasing $\dot\theta_{set_1}$, the damped control logic fired longer pulses which allowed the propellant consumption to be reduced due to the increase in I_{sp}. The pulsing frequency of the damped controller is also reduced below that of the pseudo-rate controller since longer pulses are delivered. For smaller bias torques, the propellant consumption is smaller and the reduction in propellant consumption realized by using the damped control logic is smaller.

C. Destabilizing Disturbance Torques

Fig. 7 presents a performance comparison for an unbiased destabilizing disturbance torque.

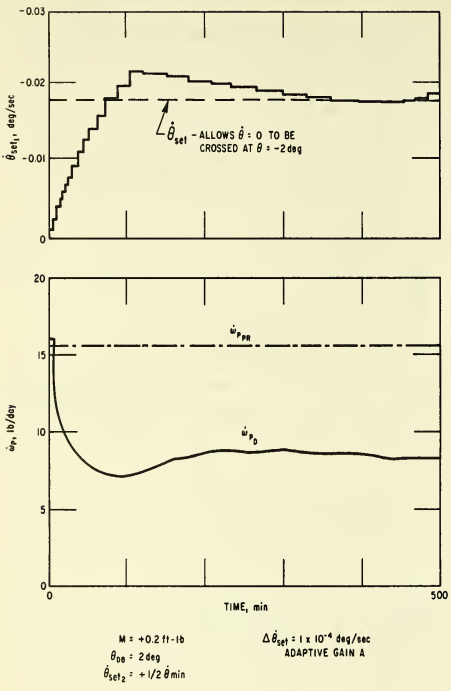


Figure 6. Constant Bias Disturbance Torque Performance

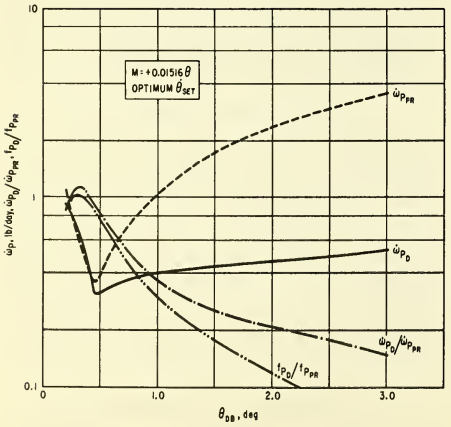


Figure 7. Unbiased Destabilizing Disturbance Torque Performance

The performance of the two controllers shown as a function of angular deadband is comparable for small θ_{DB} (small disturbance torques) because the performance is restricted by the minimum-impulse bit. Increasing the attitude deadband (increasing disturbance torque) increases the effectiveness of the damped controller for deadbands greater than the attitude at which the disturbance torque asymptote equals one-half the minimum-impulse bit rate. The dip in the $\dot{\omega}_P$ curves of Fig. 7 represents the attitude region where the disturbance torque asymptote is equal to one-half the minimum-impulse rate, and this condition ideally results in zero propellant consumption since the vehicle would drift to the origin of the phase-plane. The performance improvement with increasing deadband is sizable, and, at a 3 deg deadband, the damped controller would use only 14.9% of the propellant used by the pseudo-rate controller. Most of the increase in performance arises because the vehicle's state is allowed to approach the hyperbola asymptote much closer than the pseudo-rate controller will allow; however, two other effects influence the performance at the larger deadbands. The pulse-width increases as the disturbance torque increases (deadband increases) in the damped controller since the desired rate cutoff setting increases. The performance is improved when the pulse-width increases because of the increase in specific impulse and the decrease in impulse error. The pulsing frequency comparison, $f_{P_D}/f_{P_{PR}}$ curve of Fig. 7, follows the same trend as the propellant consumption comparisons, except that greater performance is shown in the damped controller when large disturbance torques are present. At a 3 deg deadband the damped controller would accumulate only 6.3% of the total pulses used by the pseudo-rate controller. This performance gain is due to the fact that, even though a larger amount of propellant is called for at the deadband crossing when the disturbance torque is increased, the propellant is delivered in a single pulse with the damped controller.

Fig. 8 presents the performance of the two controllers for a destabilizing disturbance torque with a large zero-torque-axis offset. Figs. 7 and 8 show that as the deadband increases (disturbance torque magnitude increases), the effectiveness of the damped controller increases. The best performance advantage is seen to exist in the region of greatest propellant consumption and pulsing frequency.

The adaptive logic contained in the damped controller will always attempt to converge to the minimum propellant expenditure trajectory when a destabilizing disturbance torque is present. When the zero-torque axis is within the deadband, the minimum propellant expenditure trajectory results from rate cutoff settings which correspond to the hyperbola asymptote crossings at the deadband. When the zero-torque axis is outside the deadband, the minimum propellant expenditure trajectory results from a rate cutoff setting which allows the zero-rate axis to be crossed just inside the opposite deadband. The adaptive performance is shown for one zero-torque-axis location in Fig. 9. The worst case (adapting over the widest range of rate cutoff variations) was taken for the adaptive range, i.e., the initial rate cutoff was set at $\theta_{min}/2$. The rate cutoff will normally be limited such that it cannot be

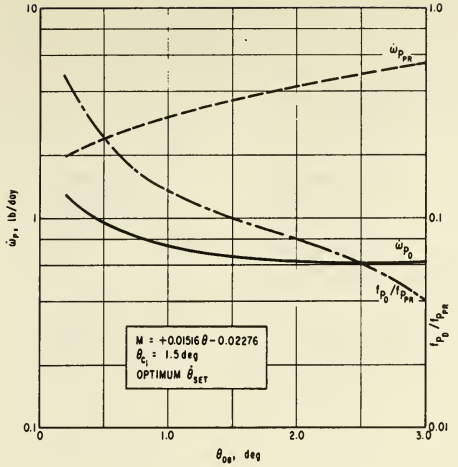


Figure 8. Biased Destabilizing Disturbance Torque Performance

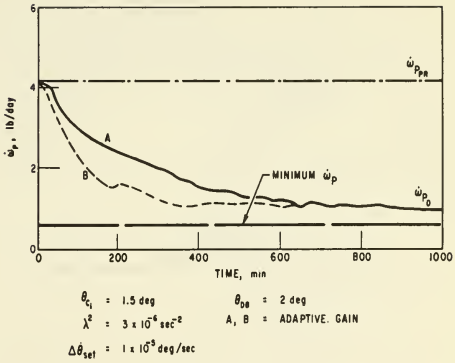


Figure 9. Adaptive Performance, Destabilizing Disturbance Torque

decreased below $\dot{\theta}_{min}/2$, and the rate cutoff corresponding to the largest value obtained from the hyperbola asymptote crossing at the deadband will normally set the upper limit. In a typical elliptical orbit where this range is possible, the disturbance torques will slowly vary through this range and the adaptive compensation will try to adapt to this change. The A and B damped curves of Fig. 9 refer to the respective adaptive gains employed, which were described in the discussion of the damped controller mechanization.

The fixed incremental change ($\Delta\dot{\theta}_{set}$) and the gains (K_1' and K_2') of the rate cutoff setting will determine how fast and how close the rate cutoff setting approaches the optimal setting. A small $\Delta\dot{\theta}_{set}$ with a large nonlinear adaptive gain (function of the deadband counter signal) is desired in order to realize fast convergence and convergence close to the optimum setting. Better performance is expected with this type of gain even though, after a period of time, the convergence is not as good as it would be if a small nonlinear gain were employed. This type of gain is recommended since the disturbance torques are time-variant. When a specific mission is chosen, the adaptive gain may be tailored to the expected range and time-variant nature of the disturbance torques.

Variations in the rate signal noise characteristics and rate signal resolution are shown in Fig. 10 which presents the performance for a wide (2 deg) and a narrow (0.5 deg) deadband. The performance is less affected, i.e., percent propellant change is less, when a wide deadband is employed since the nominal rate cutoff value is much larger. The effects of these error sources are more significant when destabilizing torques are present since a specific rate cutoff point is desired and performance is directly a function of the actual rate cutoff.

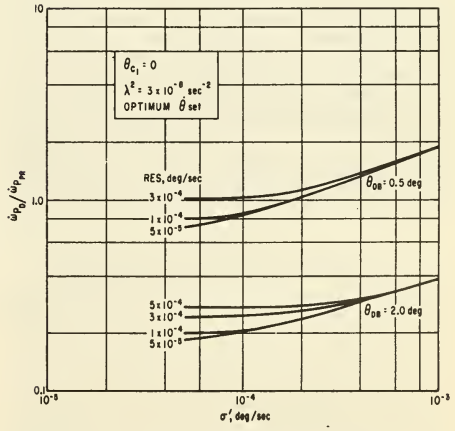


Figure 10. Effect of Rate Signal Noise and Resolution Variations, Destabilizing Disturbance Torque

D. Stabilizing Disturbance Torques

Fig. 11 presents the performance of the two controllers for a stabilizing disturbance torque with a large zero-torque-axis offset. The

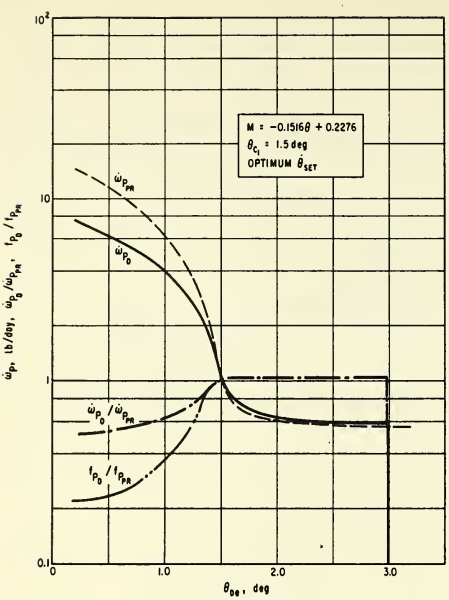


Figure 11. Stabilizing Disturbance Torque Performance

$\dot{\omega}_{P_D}/\dot{\omega}_{PPR}$ curve of Fig. 11 shows that performance can be improved when the zero-torque axis is outside the deadband, i.e., $\theta_{DB} < 1.5$ deg, which is the region of greatest propellant consumption. For 1.5 deg $< \theta_{DB} < 2\theta_{c_i}$, the performance is comparable with the pseudo-rate controller. For $\theta_{DB} > 2\theta_{c_i}$, the vehicle motions are damped within the deadband limits for the damped controller. The pulsing frequency comparison follows the same trend as the propellant consumption comparison, except that relatively better performance is shown by the damped controller for $\theta_{DB} < 1.5$ deg.

For $\theta_{DB} \geq 3$ deg in Fig. 11, the vehicle motion is damped to a trajectory which does not cross the deadband after a few pulses when the damped controller is employed. The performance for this case is governed by the time required by the adaptive logic to damp the vehicle motion, plus the time-variant nature of the disturbance torques and movement of the reference axis relative to the zero-torque axis. For constant disturbance-torque coefficients, the time required to damp the vehicle motion within the deadband is a function of the rate cutoff setting, angular deadband, adaptive gain schedule, errors, and disturbance-torque magnitude. The number of control cycles required to damp the vehicle's motion within the deadband as a function of angular deadband was obtained for the worst adaptive case, i.e., when the initial rate cutoff setting is at the smallest value. For a large disturbance torque, $\lambda^2 = -3 \times 10^{-5}$ sec^{-2}, the vehicle was damped within the deadband after the first pulse for all the variations considered.

Fig. 12A presents the performance for a small disturbance torque, $\lambda^2 = -7 \times 10^{-7}$ sec^{-2}, when the zero-position sensor error, θ_ϵ, is varied. The associated time and number of pulses required to damp the vehicle showed only

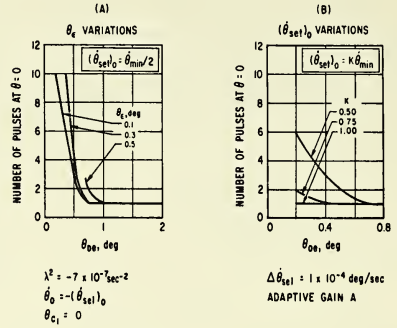


Figure 12. Number of Control Cycles Required to Damp Vehicle Motion Within the Deadband

a minor variation with position sensor errors; however, the parameters increase drastically as the deadband is reduced below 1 deg. This is due to the rate at $\theta = 0$ being less than $\dot{\theta}_{min}$. When $\dot{\theta}_{min}$ is delivered at $\theta = 0$, the residual rate (function of $\dot{\theta}_{set}$) is great enough that the same deadband is crossed for small deadbands. For deadbands greater than 1 deg, the residual rate is small enough so that the deadband is not crossed again. The effect of position sensor errors are not as important as the magnitude of the deadband unless the position error is nearly equal to the deadband limit.

Fig. 12B presents the performance as a function of the initial rate cutoff setting. The time and pulses required to damp the vehicle within the deadband reduce significantly as $(\dot{\theta}_{set})_0$ is increased. It should be mentioned that a different error distribution sequence was used for this curve, which accounts for the difference in the values shown on Figs. 12A and 12B.

The performance for rate signal resolution (1×10^{-4} to 3×10^{-4} deg/sec) and standard deviation of the rate signal noise (1×10^{-4} to 1×10^{-3} deg/sec) variations fell within the 0.1 and 0.3 deg θ_ϵ curves of Fig. 12A. Adaptive rate setting increment, $\Delta\dot{\theta}_{set}$, variations showed negligible changes in performance. The performance was more a function of the random error sequence, angular deadband, and initial rate cutoff setting than of the adaptive compensation or system errors.

When stabilizing disturbance torques with small zero-torque-axis offsets are present, the damped controller within a short period of time will damp the vehicle's motion such that another

deadband is not crossed. The time and number of pulses required to damp the vehicle's motion is mainly a function of the existing rate cutoff setting. The controller damps the vehicle's motion in the presence of realistic propulsive and sensing errors.

The performance trend for a large biased stabilizing disturbance torque is shown in Fig. 13 as a function of the rate cutoff setting without the adaptive compensation. The actual propellant

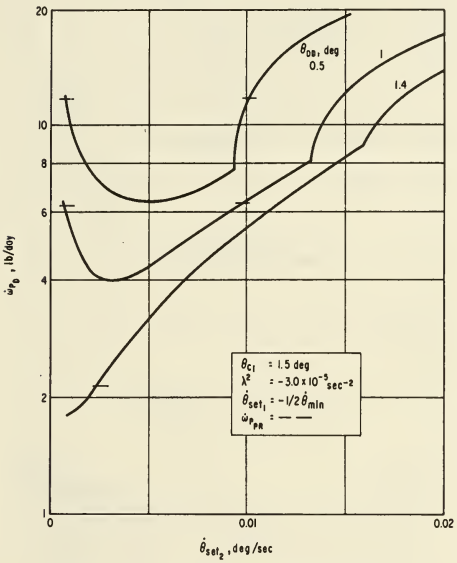


Figure 13. Performance For Biased Stabilizing Disturbance Torque

consumption and performance change of the damped controller are small for small stabilizing disturbance torques. The performance improvement at small values of $\dot{\theta}_{set_2}$ is due to the increase in I_{sp} as longer pulses are permitted. The sharp rise in propellant consumption is caused by the vehicle crossing the opposite deadband. The degraded performance of the damped controller when compared with the pseudo-rate controller for the larger deadbands is of concern, but the range of $\dot{\theta}_{set2}$ presented in Fig. 13 is larger than would normally be expected in a typical mission.

The adaptive logic for biased stabilizing torques provides convergence to either the $\dot{\theta}_{set}$ which allows the deadband furthest away from the zero-torque axis to be crossed near zero rate or to the upper limit of the rate cutoff setting, whichever is smaller. Fig. 14 presents the adaptive performance corresponding to the θ_{DB} = 0.5 deg curve of Fig. 13. The damped propellant consumption curve of Fig. 14 shows that the actual propellant consumption (9.0 lb/day) is

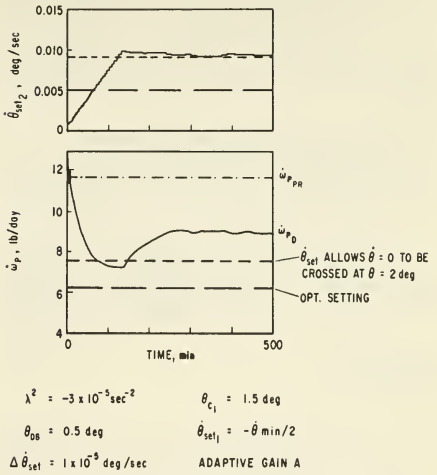


Figure 14. Adaptive Performance, Biased Stabilizing Disturbance Torque

1.35 lb/day higher than the expected value. Fig. 13 shows that the $\dot{\theta}_{set}$ for the actual performance is 0.00942 deg/sec compared to 0.0093 deg/sec for the expected value. The large difference in propellant consumption is due to the occasional crossing of the deadband farthest from the zero-torque axis and the sharp rise in propellant consumption when this deadband is crossed (Fig. 13). For biased stabilizing torques, the adaptive logic must be limited according to the range of disturbance torques expected. For example, the θ_{DB} = 1 deg curve of Fig. 13 shows that, for propellant consumption reduction, $\dot{\theta}_{set_2}$ must not exceed 0.0097 deg/sec. Fig. 13 also shows that as propellant consumption increases (θ_{DB} reduces), the propellant reduction realized by using the damped controller increases.

E. Minimum-Impulse Variations

The effects of varying the minimum pulse-width (Δt_m) or minimum-impulse bit have been described throughout this paper but have not been presented as such. The effects of varying the minimum-impulse bit can be viewed as varying $\dot{\theta}_{set}$ in the damped controller. As seen in this paper, arbitrarily increasing or decreasing a fixed $\dot{\theta}_{set}$ (minimum pulse-width) will not necessarily increase the performance since the optimum $\dot{\theta}_{set}$ varies with the disturbance torque. The adaptive logic of the damped controller seeks this optimum $\dot{\theta}_{set}$.

F. Acquisition Characteristics

The damped controller due to the rate cutoff characteristic, is relatively independent of the initial conditions after the first pulse, assuming the rate cutoff is near the proper setting. However, a pseudo-rate controller which is designed for proper limit-cycle operation goes through an

extended acquisition phase due to the multiple minimum-impulse characteristic of the system. The pseudo-rate controller uses more propellant, requires more time, and has a greater attitude excursion than does the damped controller. Due to the extremely poor performance characteristics of the pseudo-rate controller, acquisition is usually obtained from just the attitude and rate signal, thereby bypassing the pseudo-rate control logic.

Three-Axis Dynamic Simulation Results

Performance of the damped controller is compared with that of the minimum-impulse controller presented in Fig. 15. A polar orbit

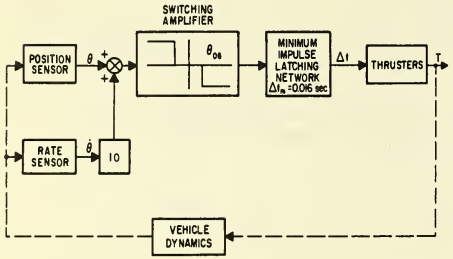


Figure 15. Minimum-Impulse Control Logic

was simulated with a perigee altitude of 90 n. mi. located 30 deg above the Equator and an apogee altitude of 200 n. mi. The spacecraft is referenced to the local vertical and orbit plane. The simulation includes gravity-gradient and aerodynamic torques. The atmosphere is assumed to rotate with the earth. The aerodynamic reference area was 78 ft^2 and the nominal center of pressure location was 17 ft behind the center of mass on the longitudinal (roll) axis. The center of mass was offset from the nominal position - 1/2 ft along each axis. The pitch and yaw control torques were 250 ft-lb and the moment arms were 7.5 ft. The roll axis values were one-tenth of the pitch and yaw values. The moments of inertia were 148,000 slug-ft^2 about the pitch and yaw axes and 6,650 slug-ft^2 about the roll axis. Although sensor and propulsive errors were not considered in this simulation, the minimum integration step size was one millisecond. The maximum integration step size was 10 secs. The simulation was started at perigee and the total mission time was 8500 sec (≈1-1/2 orbits). The attitude deadbands were ±2 deg in each axis, and all initial attitudes and rates were taken as zero.

Phase-plane trajectories for this simulation are presented in Figs. 16 through 21. Adaptive gain C, described earlier, was used for the damped controller with $\Delta\dot{\theta}_{set} = 1 \times 10^{-4}$ deg/sec. The rate cutoff setting variations plus the minimum and maximum allowable values are shown in these figures. The sharp discontinuity shown at time 5240 sec in Fig. 19 is due to the rapid movement of the zero-torque axis. The discontinuity occurs near perigee and the disturbance torques are stabilizing throughout the discontinuity. The position of the zero-torque axis is -0.77, -0.85, and -1.14 deg at respective times of 5200, 5240, and 5500 sec. The discontinuities in Fig. 21, e.g., at time 167 sec, are due to engine firings about other axes. The discontinuities are evident in the roll axis because the pitch and yaw engines are not pure couples while the roll engines are. When the pitch and yaw engines are fired, a torque is felt about the roll axis due to center of mass offsets.

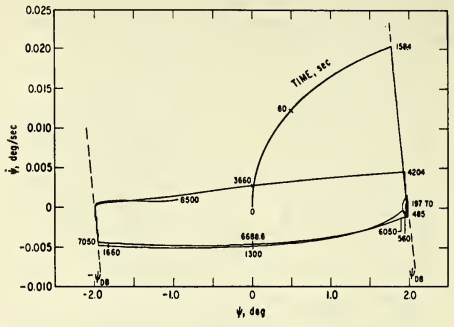


Figure 16. Yaw Phase-Plane, Minimum-Impulse Controller

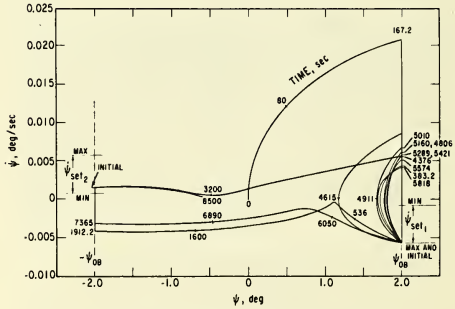


Figure 17. Yaw Phase-Plane, Damped Controller

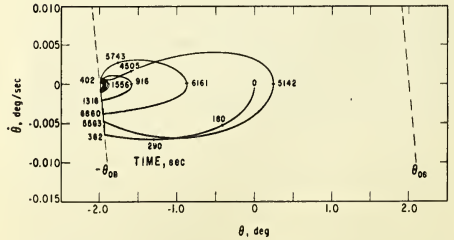


Figure 18. Pitch Phase-Plane, Minimum-Impulse Controller

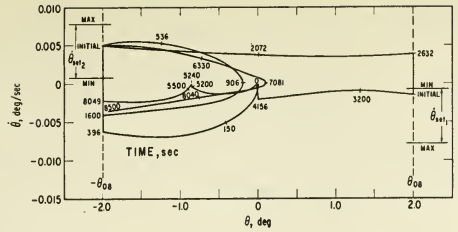


Figure 19. Pitch Phase-Plane, Damped Controller

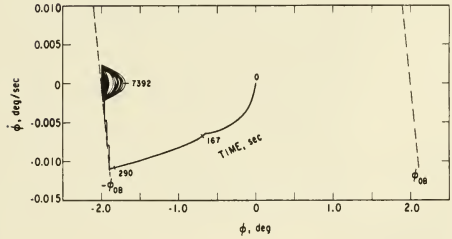


Figure 20. Roll Phase-Plane, Minimum-Impulse Controller

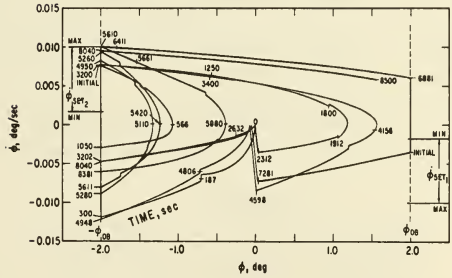


Figure 21. Roll Phase-Plane, Damped Controller

The performance is compared in Table 1. Comparing the performance of the damped controller with the minimum-impulse controller yields a relative propellant consumption of

$$\dot{\omega}_{P_D}/\dot{\omega}_{P_{MI}} = 0.49$$

and a relative pulsing frequency of

$$f_{P_D}/f_{P_{MI}} = 0.20$$

Even though the initial rate cutoff settings and adaptive gains of the damped control logic were not optimized for these results, the results point out the advantage of employing the damped control logic. The performance for a postulated worst case, i.e., initial rate cutoff settings at the minimum allowable values, was also obtained. Using these initial rate cutoff values and raising the maximum rate cutoff values of the yaw and roll axes to 0.01 and 0.0125 deg/sec, respectively, yielded:

$$\dot{\omega}_{P_D}/\dot{\omega}_{P_{MI}} = 0.63, \text{ and } f_{P_D}/f_{P_{MI}} = 0.32$$

Therefore, even in the postulated worst case, the performance of the damped controller is greatly improved over that of a minimum-impulse controller. Although the performance is based on only 1-1/2 orbits, the phase-plane trajectories and therefore performance tend to repeat each orbit. This is best illustrated in Fig. 16 by comparing the duplication of the trajectories near 1300 and 6688.6 sec. The time difference is one orbital period (5388.6 sec). Therefore, the performance results are more representative of steady-state values than the 1-1/2 orbit results would tend to indicate.

IV. Conclusions

The data presented herein show that the damped controller's performance in comparison with a minimum-impulse controller is improved in regions of large disturbance torques (largest propellant consumption and greatest pulsing frequency), and is similar in regions of small disturbance torques. In realizing the full performance potential of a damped controller, good rate signal behavior is of primary importance. The performance advantage of the damped controller is insensitive to small (relative to the deadband) movements of the zero-torque axis. Since the vehicle, in general, spends more time near zero attitude when the damped controller is employed, the disturbance torque magnitude is, in

Axis and Controller / Parameter	ψ		θ		φ		Total Values	
	MI	D	MI	D	MI	D	MI	D
$\dot{\omega}_P$ (lb/day)	4.0	1.9	1.2	0.5	2.2	1.2	7.4	3.6
f_P (pulses/day)	813	122	254	51	447	122	1514	295

NOTE: MI - Minimum impulse controller
D - Damped controller

Table 1. Performance Comparison

many cases, reduced due to the attitude-dependent nature of the disturbance torques. This in itself would tend to reduce the propellant consumption. Also, the vehicle attitude accuracy is better (in a statistical sense) with the damped controller, since the vehicle spends more time near zero attitude. Results of the three-axis dynamic simulation indicate that the damped controller uses only one-half the propellant and only one-fifth of the pulses used by a minimum-impulse controller for a typical low-altitude, elliptical orbit.

The damped controller is more efficient and economical than present minimum-impulse controllers; more efficient because the engines are fired less frequently and for the maximum warranted duration which results in higher I_{sp} and lower impulse errors; and more economical because of the reduced propellant consumption. System performance degradation of the damped controller is a function of the propulsion errors, sensor errors, and the adaptability of the system. In all present minimum-impulse attitude controllers, such as the pseudo-rate controller, the control mechanization itself constrains the system performance and a deviation from optimal performance is inherent.

The improved performance of the damped controller shown for a hypergolic-bipropellant system, may be easily realized using available hardware and without unduly complicating the control system, since only additional logic and a rate signal are employed.

List of Symbols

f_p = pulsing frequency, pulses/day

I = moment of inertia, slug-ft^2

I_{sp} = specific impulse, sec

K = constant

K_A, K_B = deadband counter signals in adaptive compensation

K_o = attitude dependent disturbance torque coefficient, ft-lb/deg

K_1 = constant disturbance torque, ft-lb

K_3 = rate gain for damped controller equal to I/T, sec^2

K_4 = pseudo-rate gain, deg

K_1', K_2' = gains for adaptive compensation proportional to deadband counter signal

M = disturbance torque magnitude, ft-lb

RES = rate signal resolution, deg/sec

T = thruster torque, ft-lb

X_A, X_B = thresholds used in adaptive compensation

Δ = incremental change

Δt = thrust pulse-width, sec

Δt_m = minimum thrust pulse-width, sec

$\Delta\dot{\theta}_{set}$ = fixed incremental change for adaptive logic, deg/sec

ϵ = a small finite magnitude

θ = body pitch attitude or general attitude from reference axis, deg

θ_{c_i} = zero-disturbance-torque axis offset from reference axis equal to $-\frac{K_1}{K_o}$, deg

θ_ϵ = zero position attitude error, deg

$\dot{\theta}_{min}$ = minimum-impulse bit angular rate increment, deg/sec

$\dot{\theta}_{set_i}$, $\dot{\phi}_{set_i}$, $\dot{\psi}_{set_i}$ = rate cutoff setting for damped controller, deg/sec

λ = stabilizing or destabilizing torque parameter equal to $\sqrt{\frac{|\partial M/\partial\theta|}{I}} = \sqrt{\frac{|K_o|}{I}}$, sec^{-1}

σ = standard deviation of impulse errors, percent of commanded impulse

σ' = standard deviation of rate signal noise, deg/sec

τ = time constant, sec

ϕ, ψ = body roll and yaw attitude from reference axis, deg

$\dot{\omega}_p$ = propellant consumption rate, lb/day

$|( )|$ = absolute value of parameter

$d( )$ = differential operator

Subscripts:

D = damped controller

DB = attitude deadband equal to $\pm\left|\frac{\text{total allowable attitude range}}{2}\right|$

MI = minimum-impulse controller

PR = pseudo-rate controller

O = initial condition

1 = most positive attitude deadband limit

2 = most negative attitude deadband limit

References

1. Dahl, P. R., Aldrich, G. T., and Herman, L. K., "Limit Cycles in Reaction Jet Attitude Control Systems Subject to External Torques," ARS Paper 1920-61, 1961.
2. Englehart, W. C., "Performance Characteristics of a 'Damped' Mass Expulsion Attitude Control System Subjected to External Disturbances and System Errors," Report No. TR-1001(2307)-7, Aerospace Corp., 1967.
3. Chobotov, V. A., "Digital Simulation for Satellite Attitude Determination and Control," Report No. TOR-669(6540)-1, Aerospace Corp., 1965.
4. Nicklas, J. C. and Vivian, H. E., "Derived-Rate Increment Stabilization: Its Application to the Attitude Control Problem," JACC Paper 61-JAC-9, 1961.

-- NOTES --

-- NOTES --

No. 67-536

DETERMINATION OF THE ATTITUDE OF A SPINNING ROCKET UNDER THRUST WITH STATISTICALLY VARIED INPUTS

by

TADEUSZ PAPIS and CLARK W. BULLARD
Douglas Aircraft Company
Santa Monica, California

AIAA Paper
No. 67-536

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.02, 7.03, 7.04

-- NOTES --

DETERMINATION OF THE ATTITUDE OF A SPINNING ROCKET UNDER THRUST WITH STATISTICALLY VARIED INPUTS*

Tadeusz Papis
Senior Engineer/Scientist, Missile & Space Systems Division
Douglas Aircraft Company, Santa Monica, California

Clark W. Bullard
Engineer/Scientist, Missile & Space Systems Division
Douglas Aircraft Company, Santa Monica, California

Abstract

Attitude dispersion of a spin-stabilized rocket in vacuo is caused by a large number of primary error sources traceable to manufacturing and assembly tolerances, which are subject to statistical distribution in magnitude and are random in direction. This paper presents the equations of motion of a rocket under thrust with jet damping, variable mass and momentum effects, and all conceivable angular disturbances taken into account. The disturbances are expressed in terms of primary error sources. The solutions to these equations are evolved in such a form that their numerical integrations have to be performed once only regardless of the number of samples considered; the final results are obtained by multiplying the results of numerical integrations by coefficients determinable algebraically from statistically varying error sources. A computer program written on this basis proved to be a very useful and efficient design tool and an excellent means for investigating rocket angular motions phenomena. The results of its application to the Delta third stage vehicle indicate the need for substantial revision of the commonly held views on the relative roles played by various disturbances and jet damping during thrust.

Introduction

The attitude changes of a spinning rocket are caused by dynamic unbalances, external moments due to thrust, and perturbing impulses, all of which originate from a number of primary error sources, which are nearly always identifiable as manufacturing and assembly tolerances. The total attitude error is a vector sum of all component errors, each of which has its own source. The obvious, and for the most part, analytically determinable relationship between production tolerances and attitude dispersion has many important implications of a practical nature. Correct determination of angular motions of a space vehicle under thrust is important from the standpoint of orbital accuracy, and the ability to determine the magnitude of a component attitude error, in both absolute and relative terms, is often of a vital practical necessity because of the high cost of tight production tolerances, which tends to increase rapidly with larger payloads.

Nearly all of the error sources are subject to statistical distribution in magnitude and are random in direction. Furthermore, angular motions of a space vehicle between the spin-up and payload separation go through several distinct phases. Motion during each successive phase is very sensitive to the terminal time of the former phase, which is also subject to statistical variations within some limits. Also, given a full set of manufacturing tolerances of a configuration, it is virtually impossible to predict what combination of the inaccuracies would result in a maximum dispersion during the burning phase, even if the effects of the former phase motion were neglected. When the latter are included, the uncertainty is further compounded. The so-called "worst case" analyses, even if obtained by means of a six-dimensional trajectory program, are no more reliable than the guesses that lead to a particular selection of inputs. Consequently, a meaningful analysis of ensuing angular motions and realistic determination of the total and component errors requires inclusion of these statistical phenomena. This can be accomplished by an application of a Monte Carlo technique in which the magnitudes and directions of the statistically varying error sources (viz., misalignments, eccentricities, tip-off impulse, etc.) and timing differentials are randomly selected from appropriate distributions and are used as inputs to the equations to be solved. The process is repeated many times, and the results are statistically analyzed.

Once some important details concerning correct and efficient conversion of the error sources into a disturbing moment have been resolved in a manner which is consistent for all phases of motion, the statistical approach lends itself readily to non-thrusting phases, in which the equations of motion yield simple closed form solutions. However, mass and momentum transfer effects during burning limit the solution of the equations of motion under thrust to the use of numerical methods only. Because they are time-consuming, numerical methods are not generally suited for a Monte Carlo-type analysis. In the present paper, this difficulty is overcome by formulating solutions to the equations of motion in such a form that numerical integration need be performed for one sample only. The final results for each sample are obtained, with no loss of accuracy

*This paper is based on work accomplished by the Douglas Missile and Space Systems Division under Contract NAS7-380 for the Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland.

The authors wish to express their appreciation to Mr. J. W. Winchell, who, in his leading role in the spin balance analysis program, provided indispensable and sympathetic support for their efforts and constructive criticism in preparation of this paper; and to Messrs. D. L. Mootchnik and T. K. Oswald for their valuable comments and suggestions.

67-536

whatsoever, by multiplying the results of the numerical integration by coefficients determined by simple analytical functions of statistically derived inputs.

The present paper is a by-product of the authors' participation in a Douglas Aircraft Company study effort with the objective of revising the spin-balancing procedures and establishing analytical criteria for the assembly and manufacturing tolerances. The effort coincided with the introduction of the solid propellant TE-364 motor into a third stage of the Delta vehicle. The detailed analysis presented here evolved around that particular motor, which is somewhat unusual in that the center of gravity (c.g.) position of the motor and its radii of inertia remain very nearly constant throughout the entire period of burning. While this feature significantly simplifies the algebra and calculus of the problem, the fundamental approach used in the present analysis is fully applicable, as will be shown, to any spin-stabilized third stage rocket with any solid propellant motor.

The System Under Consideration

The system that shall be considered in detail is depicted schematically in figure 1.

THE SYSTEM UNDER CONSIDERATION

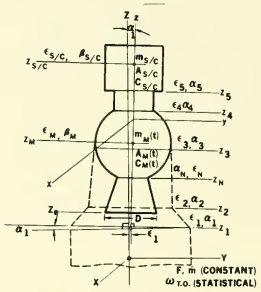


FIGURE 1

The third stage during burning is shown in bold outline. The broken outline indicates the motor attach fitting, the spin table, and the second stage, which are jettisoned prior to ignition. Spin is produced by the jet action of a number of small rockets distributed uniformly along the circular periphery of the table, which is free to rotate on bearings on the top face of the second stage.
Thus the perpendicular to the bottom plane of the spin table, which runs through its center, constitutes the nominal spin axis and the nominal axis of symmetry of the third stage. This is the z axis of a body coordinate system xyz; the plane of the x-y axes moves longitudinally so as to be always through the c.g. of the third stage. Since the attitude error of the third stage is measured with respect to the attitude held by the second stage just prior to third stage separation, the perpendicular to the top face of the second stage, which runs through its nominal center, is taken as the Z axis of the coordinates XYZ. The orientation of the Y-Z pair of axes is such that at the onset of spin y-z and Y-Z (and, hence, x-z and X-Z) are coplanar. At that time, the planes of x-y and X-Y are inclined to each other at an angle α_1. The subscripted letters z_1 to z_5 denote the stations of the mating planes 1 to 5, while α_1 to α_5 and ϵ_1 to ϵ_5 are the assembly misalignments and eccentricities at these planes. Manufacturing mass unbalances, as usually defined by appropriate specifications, are described by misalignments $\beta_{S/C}$ and β_M and by eccentricities $\epsilon_{S/C}$ and ϵ_M, while α_N and ϵ_N are nozzle misalignments and eccentricities, respectively. All of these and the angular rate $\omega_{T.O.}$, which is due to tip-off impulse applied at the third stage ignition, constitute the primary attitude error sources, which are random in direction and statistically varying in magnitude within the limits prescribed by drawings and/or specifications. Static and dynamic unbalances due to appendages (not shown on the diagram) also may be present; these, as a rule, will be fixed both in magnitude and in direction.

Since the spacecraft and the motor account for the bulk of the third stage weight (over 99 per cent at ignition, and about 97.5 per cent at burnout), the stage constitutes a dumbbell consisting of two masses $m_{S/C}$ (spacecraft plus the attach fitting) and m_M (motor) with centers of gravity at $z_{S/C}$ and z_M and with pitch (yaw) and roll inertias $A_{S/C}$, A_M, $C_{S/C}$, and C_M, respectively. Mass properties of the spacecraft remain constant throughout. For all practical purposes, the c.g. of the spherical motor also remains constant. Motor inertias are linear functions of its mass, and the mass flow $\dot{m}$ is constant. Other inputs to the problem are: thrust, F; nozzle attachment and exit stations , z_N and z_e; and nozzle diameter, D.

Euler's Equations of Motion

The system having been defined, one can proceed with the development of the equations of angular motion, taking into consideration all the pertinent characteristics of the system. From the fundamental relationship in body axes,[1] viz., $\overline{M} = \dot{\overline{h}} + \overline{\omega} \times \overline{h}$ + rate of angular momentum transfer, one obtains the well-known equations for a symmetric rocket under thrust:

$$\left.\begin{aligned} M_x &= A\dot{\omega}_x + \dot{A}\omega_x + (C-A).\omega_x.\omega_z + \Delta\dot{h}_x \\ M_y &= A\dot{\omega}_y + \dot{A}\omega_y + (A-C).\omega_y.\omega_z + \Delta\dot{h}_y \\ M_z &= C\dot{\omega}_z + \dot{C}\omega_z + \Delta\dot{h}_z \end{aligned}\right\} \quad (1)$$

where $\Delta\dot{h}_x$, $\Delta\dot{h}_y$, and $\Delta\dot{h}_z$ are the components of the rate of momentum transfer along the x, y, and z axes. Equations (1) are rigorously valid as long as the effective system asymmetry resulting from the applied moments M_x and M_y remains small. Other departures from strict validity may arise from detailed interpretation of A and components of $\Delta\dot{h}$ only. In the present analysis, this interpretation is based upon the following usual assumptions:

(1) The contribution of burned fuel mass inside the system to the moments of inertia of the system is negligible.

(2) The mass flow is constant and uniform across the exit area.

(3) Viscosity effects are negligible.

From the second assumption, it follows that $\Delta\dot{h}_x = \dot{m}(\ell_e^2 + \frac{R^2}{4})_x$ and $\Delta\dot{h}_y = \dot{m}(\ell_e^2 + \frac{R^2}{4})_y$, where $\frac{R^2}{4}$ is the square of the radius of gyration of the exit area, R being the exit radius. Unless the pattern of burning is such as to form internal vanes capable of producing Coriolis forces in transverse planes, assumption 3 leads to $\Delta\dot{h}_z = -\dot{C}\omega_z$; this and the condition $M_z = 0$ lead to $\omega_z = n$ (a constant).

As the result of the foregoing, Euler equations (1) stated in complex notation become

$$M_{xy} = A\dot{\omega}_{xy} + \dot{A}\omega_{xy} - \dot{m}\ell^2\omega_{xy} - i(C-A).n.\omega_{xy} \quad (2)$$

where

$$\ell^2 = \ell_e^2 + \frac{R^2}{4}$$

with

$$\ell_e = c.g._{STAGE} - z_e \quad (3)$$

and $i = \sqrt{-1}$.

Equations (2) are of the form

$$\dot{\omega}_{xy} + P(t).\omega_{xy} = Q(t) \quad (4)$$

where

$$P(t) = \frac{\dot{A} - \dot{m}\ell^2}{A} - i\frac{C-A}{A}.n \quad (5)$$

and

$$Q(t) = \frac{M_{xy}}{A} \quad (6)$$

Solution of (4) is of the form

$$\omega_{xy}(t) = e^{-\int_{t_o}^{t} P.dt}\left[\int_{t_o}^{t} Q.e^{\int_{t_o}^{t} P.dt}.dt + \omega_{xy}(t_o)\right] \quad (7)$$

where t_o denotes the time at liftoff (ignition) and ω_{xy} is the complex pitch and yaw rate in body axes.

Since the mass of the system m is a linear function of time

$$m = m(t_o) + \dot{m}(t-t_o), \quad (8)$$

equation (7) can be written in terms of m:

$$\omega_{xy}(m) = \quad (9)$$

$$e^{\frac{1}{\dot{m}}\int_{m_o}^{m} P(m)dm}\left[\frac{1}{\dot{m}}\int_{m_o}^{m} Q(m).e^{-\frac{1}{\dot{m}}\int_{m_o}^{m} P(m)dm}.dm + \omega_{xy}(m_o)\right]$$

where $m_o = m(t_o)$. Quantitative solution of equations (7) or (9) requires that functions Q and P be expressed explicitly as functions of time or mass. Working in terms of m will be found to be a definite advantage.

Derivation of Function Q

The major concern here will be the proper determination of the total body fixed moment M_{xy} in terms of original error sources and as a function of time (or total mass m). This often appears to be a baffling problem because of longitudinal and lateral shifts in c.g. positions, uncertainties with respect to the position of the true principal axis of the whole system, and numerous other difficulties arising from the multiplicity of the error sources, the three-dimensional character of the system of unbalances, and time-dependent changes. Once, however, some very basic relationships that govern the problem are clearly recognized, all of these difficulties are quickly reduced to readily manageable proportions.

Because of the dumbbell property of the system, the only mass unbalances of consequence are those due to the spacecraft and the motor. When spacecraft is considered separately, the only unbalances present are those due to spacecraft manufacturing tolerances and those due to a possible presence of appendages. When, on the other hand, the spacecraft is treated as a part of an assembly, then assembly eccentricities and misalignments produce an offset of the c.g. of the spacecraft with respect to the nominal spin axis of the system, resulting in a static unbalance of the spacecraft within the assembly. Furthermore, assembly misalignments cause a misalignment of the axis of symmetry of the spacecraft with respect to the same reference line, resulting in a dynamic unbalance of the spacecraft within the system. Thus the spacecraft is said to possess static and dynamic unbalances due to assembly tolerances. Precisely the same observations apply to the motor.

The dynamic unbalance of the entire stage is given by

$$(mr\ell)_{STAGE} = (mr\ell)_{s/c} + (mr\ell)_M + (mr)_{s/c}.\ell_{s/c} + (mr)_M.\ell_M \quad (10)$$

where $(mr)_{s/c}$, $(mr)_M$ and $(mr\ell)_{s/c}$, $(mr\ell)_{s/M}$ denote the total static and dynamic unbalances of the spacecraft and the motor evaluated with respect to the nominal spin axis of the system, while $\ell_{s/c}$ and ℓ_M are the distances of the respective c.g.'s from the c.g. of the entire assembly. The total static (or dynamic) unbalance of each component part of the entire system is the sum of its inherent unbalance due to manufacturing tolerances (and appendages, if present) and the unbalances resulting from assembly eccentricities and misalignments.

The equations for the total mass unbalances of the spacecraft and the motor are written in terms of the tolerances, distances, and mass properties involved; they are substituted into equation (10), and the time-dependent quantities are expressed in terms of the time-varying total mass of the stage m; the expression for the dynamic unbalance of the system thereby attains the form

$$(mr\ell)_{xy} = s_1 m + s_2 + \frac{s_3}{m} \quad (11)$$

where the complex coefficients s_1, s_2, and s_3 are simple functions of statistically varying tolerances and constant inputs concerning mass properties which do not vary with time, i.e.,

$$\left.\begin{aligned} s_1 &= (r\ell)_M \\ s_2 &= m_{s/c}\left[L\,(r_{s/c} - r_M) + (r\ell)_{s/c} - (r\ell)_M\right] \\ s_3 &= -m^2_{s/c}\,L\,(r_{s/c} - r_M) \end{aligned}\right\} \quad (12)$$

where

$$\left.\begin{aligned} (r\ell)_{s/c} &= \\ &(\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \beta_{s/c})(k^2_{s/c_R} - k^2_{s/c_P}) \\ (r\ell)_M &= (\alpha_2 + \alpha_3 + \beta_M)(k^2_{M_R} - k^2_{M_P}) \\ (r_{s/c} - r_M) &= \varepsilon_4 + \varepsilon_5 + \varepsilon_{s/c} - \varepsilon_M\,. \\ &+ \alpha_2\,(z_{s/c} - z_M) + \alpha_3\,(z_{s/c} - z_M) \\ &+ \alpha_4(z_{s/c} - z_4) + \alpha_5\,(z_{s/c} - z_5) \end{aligned}\right\} \quad (13)$$

and L is the distance between the centers of gravity of the spacecraft and the motor; k_{s/c_P}, k_{s/c_R} and and k_{M_P}, k_{M_R} are pitch and roll radii of inertia of the spacecraft and the motor, respectively. In equations (13) and wherever else they may appear, eccentricities and misalignments are expressed as directed numbers.

Proceeding along similar lines, the external moment due to thrust is expressed as

$$(M_{xy})_T = i\,s_4 + i\,\frac{s_5}{m} \quad (14)$$

where s_4 and s_5 are statistically varying complex constants given by

$$\begin{aligned} s_4 &= F(\varepsilon_N + \alpha_N \cdot d_N + r_M) \\ s_5 &= F.m_{s/c}\cdot\,(L\,.\,\alpha_N + r_{s/c} - r_M) \end{aligned} \quad (15)$$

The quantity F is thrust; d_N is the distance between the nozzle and the c.g. of the motor; $r_{s/c}$ and r_M are the offsets from the z axis of the c.g.'s of the spacecraft and the motor, respectively.

By use of the dynamic equivalence between dynamic unbalance and the body-fixed moment, viz., $M_{xy} = in^2(mr\ell)_{xy}$, where n is the spin rate, equations (11) and (12) can be combined to yield

$$M_{xy} = S_1 m + S_2 + \frac{S_3}{m} \quad (16)$$

where

$$S_1 = in^2 s_1,\ S_2 = i(n^2 s_2 + s_4),$$

$$S_3 = i(n^2 s_3 + s_5)$$

If fixed mass unbalances are present (e.g., as the result of appendages), the body-fixed moments due to these unbalances can be expressed in the form

$$(M_{xy})_{FIXED} = f_1 + \frac{f_2}{m} \quad (17)$$

where f_1 and f_2 are fixed constants formed by constant inputs, including the in^2 term and L.

The pitch moment of inertia for the stage is

$$A = A_{s/c} + A_M + m_{s/c}\ell^2_{s/c} + m_M \cdot \ell^2_M \quad (18)$$

where $\ell_{s/c}$ and ℓ_M are as formerly defined. Equation (18) reduces to

$$A = K_1 \cdot m + K_2 + \frac{K_3}{m} \quad (19)$$

where K_1, K_2 and K_3 are constant coefficients defined by constant inputs:

$$K_1 = k^2_{M_P}$$

$$K_2 = m_{s/c}(L^2 + k^2_{s/c_P} - k^2_{M_P})$$

$$K_3 = -m^2_{s/c} \cdot L^2$$

From equations (16), (17), and (19), the expression for Q(m) is finally obtained in terms of the coefficients S_1, S_2, and S_3 determined by statistically varying tolerances and fixed inputs:

$$Q(m) = \frac{S_1 m^2 + (S_2 + f_1)m + S_3 + f_2}{U(m)} \quad (20)$$

where

$$U(m) = mA(m) = K_1 m^2 + K_2 m + K_3 \quad (21)$$

Expressions Involving Function P

Since

$$\dot{A} = \dot{m}\,\frac{dA}{dm}\ ,$$

it follows that

$$\int P(t)dt = \int\left(\frac{A' - \ell^2}{A} - i\,\frac{C - A}{A}\,n\right)dt \quad (22)$$

where A' is $\frac{dA}{dm}$.

By differentiating equation (19), expressing ℓ as a function of m, and performing the necessary algebraic manipulation, the following is obtained:

$$A' - \ell^2 = k_1 + \frac{k_2}{m} \tag{23}$$

where

$$\left.\begin{aligned} k_1 &= k_{M_P}^2 - (d_e^2 + \frac{R^2}{4}) \\ k_2 &= -2m_{s/c}.L.d_e \end{aligned}\right\} \tag{24}$$

The quantity d_e is the distance between the c.g. of the motor and the nozzle exit.

When expressed as a function of m, inertia difference becomes

$$-(C-A) = k_3 m + k_4 + \frac{k_5}{m} \tag{25}$$

where

$$\left.\begin{aligned} k_3 &= k_{M_P}^2 - k_{M_R}^2 \\ k_4 &= m_{s/c}(L^2 + k_{s/c_P}^2 - k_{M_P}^2 - k_{s/c_R}^2 + k_{M_R}^2) \\ k_5 &= -m_{s/c}^2.L^2 = K_3 \end{aligned}\right\} \tag{26}$$

Equation (22) now can be expressed in terms of m and constants:

$$\int P(t)dt = \int \frac{k_1 m + k_2}{U(m)}\,dm + i\frac{n}{\dot{m}} \int \frac{k_3 m^2 + k_4 m + k_5}{U(m)}\,dm \tag{27}$$

This is a directly integrable form. Integration between the limits t_o and t (i.e., m_o and m), leads to the following solution:

$$\int_t^t P(t)dt = \left[\xi_1 \mathrm{Log}U + \xi_2 \mathrm{Log}V + i\frac{n}{\dot{m}}(\xi_3 m + \xi_4 \mathrm{Log}U + \xi_5 \mathrm{Log}V)\right]_{m_o}^{m} \tag{28}$$

where

$$\xi_1 = \frac{k_1}{K_3}$$

$$\xi_2 = (k_2 - \frac{K_2}{2K_1}k_1)\frac{1}{\sqrt{K_2^2 - 4K_1K_3}}$$

$$\xi_3 = \frac{k_1}{K_1}$$

$$\xi_4 = \frac{k_4 - \xi_3 K_2}{2K_1}$$

$$\xi_5 = \left\{k_5 - \frac{K_2}{2K_1}k_4 - \xi_3(K_3 - \frac{K_2}{2K_1}K_2)\right\}.\frac{1}{\sqrt{K_2^2 - 4K_1K_3}}$$

and

$$V = \frac{m-p}{m-q}$$

where p and q are the roots of U(m).

Hence,

$$e^{\int_{t_o}^{t} P(t)dt} = (\frac{U}{U_o})^{\xi_1}.(\frac{V}{V_o})^{\xi_2}.e^{i(\phi-\phi_o)} \tag{29}$$

where

$$\phi = \frac{n}{\dot{m}}(\xi_3 m + \xi_4 \mathrm{Log}U + \xi_5 \mathrm{Log}V)$$

and

$$U_o = U(m_o),\quad V_o = V(m_o),\quad \phi_o = \phi(m_o)$$

Solutions for ω_{xy} and θ_{XY}

By inserting equations (20) and (29) in equation (9) and denoting, for brevity, the definite integral of(29) by W, the following expression for the pitch and yaw rates is obtained:

$$\omega_{xy} = \frac{1}{\dot{m}W}\left\{S_1\int_{m_o}^{m}\frac{W}{U}m^2dm + (S_2+f_1)\int_{m_o}^{m}\frac{W}{U}m.dm + (S_3+f_2)\int_{m_o}^{m}\frac{W}{U}dm\right\} + \frac{\omega_{xy}(m_o)}{W} \tag{30}$$

where the coefficients S_1, S_2, and S_3, and $\omega_{xy}(0)$, which may include the tip-off impulse, are the only quantities containing the statistically varying inputs. In order to obtain ω_{xy}, numerical integration of the integrals present will have to be performed once only, regardless of the number of statistical samples.

For small angular displacements, the attitude angle θ measured in inertial coordinates is given by

$$\theta_{XY} = \int \omega_{xy}\, e^{in\int dt} + c$$

which, when applied in the present coordinate system as defined, is

$$\theta_{XY}(t) = \int_{t_o}^{t} \omega_{xy}(t)\, e^{in\int dt} + \theta_{XY}(t_o) \qquad (31)$$

where t is the time measured from the onset of spin.

A change in the independent variable from t to m yields

$$\theta_{XY}(m) = \frac{e^{in\int_{o}}}{\dot{m}} \int_{m_o}^{m} \omega_{xy}(m) e^{in\frac{m-m_o}{\dot{m}}}\, dm + \theta_{XY}(m_o) \qquad (32)$$

By combining equation (30) with (31), the final expression for the attitude displacement is obtained in terms of $m = m(t)$:

$$\theta_{XY}(m) = \frac{e^{in\int_{o}}}{\dot{m}} \left[S_1 \int_{m_o}^{m} \frac{D_1}{W} e^{in\frac{m-m_o}{\dot{m}}}\, dm \right.$$

$$+ (S_2+f_1) \int_{m_o}^{m} \frac{D_2}{W} e^{in\frac{m-m_o}{\dot{m}}}\, dm$$

$$+ (S_3+f_2) \int_{m_o}^{m} \frac{D_3}{W} e^{in\frac{m-m_o}{\dot{m}}}\, dm$$

$$\left. + \int_{m_o}^{m} \frac{\omega_{XY}(m_o)}{W} e^{in\frac{m-m_o}{\dot{m}}}\, dm \right] + \theta_{XY}(m_o) \qquad (33)$$

where functions

$$D_1 = \frac{1}{\dot{m}} \int_{m_o}^{m} \frac{W}{U} m^2 dm \;,\quad D_2 = \frac{1}{\dot{m}} \int_{m_o}^{m} \frac{W}{U} m.dm \;,$$

and $$D_3 = \frac{1}{\dot{m}} \int_{m_o}^{m} \frac{W}{U}\, dm$$

are available from the former integration in equation (30). Again, evaluation of equation (33) requires one numerical integration of all the integrals, regardless of the number of statistical samples.

Extensions of the Method

Cylindrical motors, such as FW-4 used in the third stage Delta vehicle, experience considerable changes in c.g. position and radii of inertia during burning. If the motor c.g. and radii of inertia are to be treated as time-dependent variables, we express z_M, $k^2_{M_P}$, and $k^2_{M_R}$ in terms of m(t) and proceed as formerly. Any conceivable time history of $k^2_{M_R}$ and $k^2_{M_P}$ for a solid motor can be represented well within the accuracy of data by means of a quadratic in m(t). The movement of the c.g. of the motor can have only a minor effect on the c.g. position of the stage because changes in the latter are produced mainly by mass variations in the motor; the effect of the movement of the c.g. upon angular motions will be even less significant. Thus, linear approximation of $z_M(t)$ will always be more than adequate.

With these modifications, the final expression for Q(m) attains the form

$$Q(m) = \frac{S_1 m^4 + S_2 m^3 + S_3 m^2 + S_4 m + S_5}{U(m)} \qquad (34)$$

where U(m) is a quartic in m.

Similarly, function P(m) now has the general form

$$P = \dot{m}\, \frac{k_1 m^3 + k_2 m^2 + k_3 m + k_4}{U(m)}$$

$$+ in\, \frac{k_5 m^4 + k_6 m^3 + k_7 m^2 + k_8 m + k_9}{U(m)} \qquad (35)$$

which is integrable in closed form. Consequently, the final expressions for ω_{xy} and θ_{xy} each require one numerical integration only, although the number of integrals will increase.

While the foregoing extensions seem to be routine in nature, their desirability is questionable, in general. In the case of the Delta third stage vehicle, they are simply not necessary. Mass changes dominate the issues throughout, and numerous small effects of the motor c.g. movement upon the total effective asymmetry angle of the system β tend to cancel each other as well as the effect due to changes in the motor's radii in inertia. The net effect upon the time history of angular motions is too small to be considered accountable in any practical terms. Furthermore, the presence of quartics in equations (34) and (35) will place highly increased demands upon the accuracy of numerical integrations to be carried out, and an increase in the number of integrals and of the attendant statistically derived coefficients will increase the machine time. Curve fitting is never desirable.

Computer Program

On the basis of equations (30) and (33), a program has been written for the Univac 1108 computer with provisions to permit its use for two distinct purposes:

(1) as a design tool for determination of the third stage attitude error at injection of the payload into the orbit and/or for determination of allowable tolerances, their effect upon attitude error, orbital accuracy, etc.

(2) as an investigative tool for post-flight analysis work and related purposes, such as diagnosis of possible causes for attitude errors experienced during flight, determination of angular motions time history, etc.

In the first application, the program constitutes one phase of a multiphase third stage attitude analysis program. Having computed and stored the values of the integrals of the equations (30) and (33), the computer selects, by a Monte Carlo process, the error sources from appropriate statistical distributions and solves all the phases of motion. Sample size and statistical distributions can be varied. The terminal phase outputs are subjected to statistical analyses and are plotted on the Stromberg-Carlson SC-4020 graphical output device with as many errors due to separate sources singled out as desired (figure 2). What constitutes the terminal phase is, of course, a matter of the user's choice. For the thrusting phase, the output ordinarily consists of the final attitude error and the pitch and yaw rate, and the average attitude error during the phase. The outputs for non-thrusting phases may include the center of precession, coning angles, and other quantities, as required. In the second application, the graphical output consists of complete time histories of ω_{xy} and θ_{xy} during burning (figures 4 through 10). The values of integrals stored at each time point are multiplied by one set of statistically or arbitrarily* selected inputs, and the results are plotted by the machine. If desired, the jet damping and all or just some thrust effects can be suppressed. In this manner the effects of any particular factor or a group of factors upon the motion can be studied and demonstrated very effectively.

STATISTICAL ANALYSIS AUTOMATED OUTPUT (A SAMPLE)

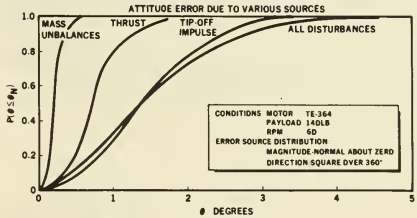


FIGURE 2

ATTITUDE ERROR VS PAYLOAD WEIGHT AT $P(\theta \leq \theta_N) = .99$

MOTOR: TE-364
(a) MASS UNBALANCES ONLY
(b) THRUST ONLY
(c) TIP-OFF IMPULSE ONLY
(d) ALL ERROR SOURCES INCLUDED

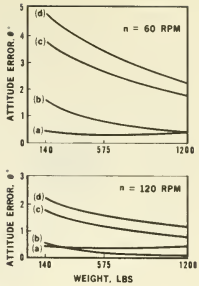


FIGURE 3

ω_{xy} DURING THRUST INITIAL ANGULAR RATE ONLY

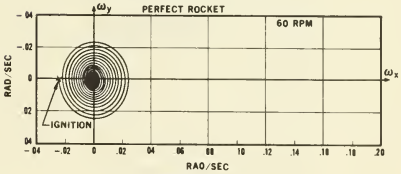


FIGURE 4

θ_{XY} DURING THRUST INITIAL ANGULAR RATE ONLY

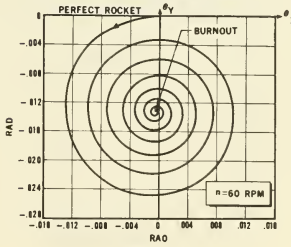


FIGURE 5

ω_{xy} FOR REAL ROCKET UNDER THRUST INITIAL ANGULAR RATE PRESENT

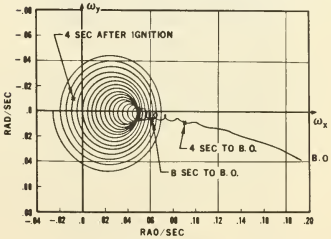


FIGURE 6

* It is noted that inputs to cases presented in figures 4 through 10 were selected arbitrarily with the intention of showing certain aspects of the motion.

67-536

ω_{xy} DURING THRUST
NO INITIAL ANGULAR RATE

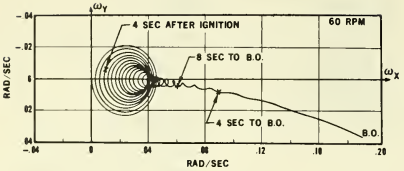


FIGURE 7

θ_{XY} FOR REAL ROCKET UNDER THRUST
INITIAL ANGULAR RATE PRESENT

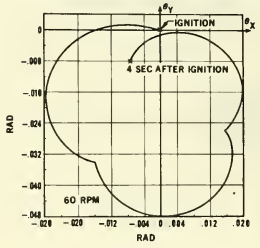


FIGURE 8

θ_{XY} FOR REAL ROCKET UNDER THRUST
NO INITIAL ANGULAR RATE

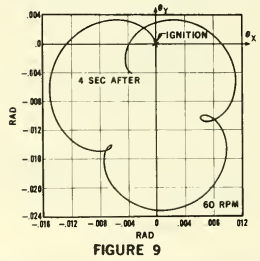


FIGURE 9

θ_{XY} FOR REAL ROCKET DURING
CLOSING STAGES OF BURNING

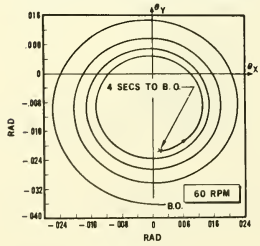


FIGURE 10

Conclusions

The program was applied to the third stage Delta vehicle, with a TE-364 motor and a wide range of payload weights, within the scope of a study aimed at determination of criteria for manufacturing tolerances and spin balancing. With 17 statistically varying inputs, the machine time necessary to obtain angular rates and displacements during the thrusting phase was 58 seconds for one sample and 1 minute and 15 seconds for 1000 samples. The need for statistical treatment of the error sources and the usefulness of the program, both as a design tool and as a means for investigating angular motions of a spinning missile under thrust, were confirmed beyond any doubt. The program outputs, an exemple of which is presented in figure 2, took the guesswork out of design decisions. As a matter of interest, note the intersection of the total attitude error curve with that of the predominating error component (in tnis case, the tip-off impulse) at approximately the 50 per cent probability level. The observed crossover reflects vectorial summation of component attitude errors determined through statistical selection of error sources. The form of statistical distribution of the tolerances is frequently a subject of considerable uncertainty and controversy. The program lent inself well to investigation of the effects of various distributions (formed around the stated tolerance limits) upon the attitude error. As expected, the maximum and/or near-maximum attitude error values were found to be almost independent of the precise form of the tolerance error distribution within the prescribed limits. The shape of the plot of the error probability versus the attitude error is, of course, strongly affected by the form of statistical distribution assumed. The graphs in figure 3 are the cross plots obtained from a number of machine outputs, such as those in figure 2. They show the variations of the component errors and the total error with changing payloads at 60 and 120 r.p.m. and at a 99 per cent probability level.

The outputs, such as those in figures 4 to 10, proved very informative with regard to the role played by various parameters of angular motions during burning. Within the limits of this paper, one can make only a few observations in the briefest of terms with regard to the meaning of these plots. Figures 4 and 5 illustrate the effect of jet damping in the hypothetical case of a perfect rocket with no mass unbalances or external moments (other than jet damping moment); in this case existence of angular motions is caused entirely by initial angular rate. While the amplitude of transverse oscillations is steadily being damped out (figure 4), the oscillations continue with a frequency that is only slightly influenced by changes in the relative spin velocity produced by variations in the inertia ratio of the body. As a result, the rocket remains in a state of pure precession (figure 5) right up to the time when the coning angle becomes vanishingly small. The position of the center of precession experiences only a slight shift in space; this shift is produced by jet damping and the cumulative effect of variations in relative spin velocity. Eventually the center of precession becomes fixed in space and coincident with the spin axis of the rocket.

Although the foregoing case is frequently discussed in connection with rocket angular motions during burning (possibly because it lends itself easily to a simple demonstration of a jet damping effect), it bears little resemblance to an actual situation obtained with real rockets, particularly during the later stages of motion, which are the most important ones from the standpoint of the attitude error subsequent to burnout. The essential features of angular motions of a real rocket under thrust are reflected in figures 6 through 10; these figures clearly show the predominant effect, increasing with time, of external moment due to thrust, and they illustrate the limitation in the role that can be played by jet damping (no matter how strong) in the presence of that moment. Oscillations in pitch and yaw rates ω_x and ω_y are, indeed, damped out (figures 6 and 7), and the originally present precessional motion with nutations (figures 8 and 9) no longer exists (figure 10). However, a continuous (if somewhat restricted and modified by jet damping) increase in the effective asymmetry angle $\beta_{xy} = \frac{M_{xy} + \text{jet damping moment}}{(C - A)n^2}$ of the system persists at an ascending rate owing to an increase in the thrust-produced moment (c.g. of the system moving rapidly up) and a decrease in the inertia difference (C - A). As the rocket attains a state of almost pure spin about the shifting axis of dynamic equilibrium (defined by β), the transverse angular rates ω_x and ω_y increase (figures 6 and 7) while the vehicle's nominal axis of symmetry follows a spiralling out path (figure 10) with an angular velocity equal to the spin rate. With the end of burning, both the moment due to thrust and jet damping will cease to exist; these events will be accompanied by a relatively big vectorial change in the effective asymmetry angle β of the system and in resumption of precessional motion, most likely of significant proportions, by the rocket. The resultant mean attitude error and the coning angle depend largely upon the phase relationship between the new angle β and the angular rate at the end of burning. The new angle β is determined by statistical distribution of numerous primary error sources; this fact provides another illustration of the overall need for a statistical approach to the problem of attitude error determination.

1. Thompson, W. T., Introduction to Space Dynamics, New York: John Wiley & Sons, Inc., 1961, p. 223.

-- NOTES --

No. 67-537

ANALYSIS OF JET ATTITUDE CONTROL SYSTEMS SUBJECT TO VARYING MAGNITUDES OF EXTERNAL DISTURBING TORQUES

by

V. E. HALOULAKOS
Douglas Aircraft Company
Santa Monica, California

AIAA Paper
No. 67-537

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.01, 7.03, 7.04

ANALYSIS OF JET ATTITUDE CONTROL SYSTEMS SUBJECT TO VARYING MAGNITUDES OF EXTERNAL DISTURBING TORQUES*

V. E. Haloulakos
Douglas Aircraft Company, Inc.
Santa Monica, California

Abstract

Behavior of on-off jet attitude control systems when subjected to varying magnitudes of external disturbing torques is investigated. Results indicate three distinct regions of operation with different limit cycle characteristics. Of particular interest is a region, termed "ultralow disturbance region," in which the system steady-state solutions are limit cycles composed of one or more double-pulse cycles and always one single-pulse cycle. Furthermore, it is shown that the number of the double-pulse cycles increases with decreasing disturbance magnitudes and that the energy consumption in this region varies in a discontinuous manner, but remains below the level corresponding to zero disturbance. When the disturbance exceeds a precisely defined critical value, the limit cycle is the single-pulse type with energy requirements varying linearly with the disturbances.

Nomenclature

α	=	angular acceleration, rad/sec^2
$\dot{\theta}$	=	angular velocity, rad/sec
θ	=	position angle, radians
δ	=	deadband angle, radians
t_r	=	start time lags (from "on" signal to full thrust), seconds
t_d	=	cut-off time lag (from "off" signal to zero thrust), seconds
t_o	=	generalized time lag ($t_o = t_d$), seconds
k	=	switching function slope (lead time), seconds
t	=	time, seconds
t_{on}	=	control "on" time, seconds
ER = 1/(1+f)	=	energy ratio
t_{off}	=	control "off" time, seconds
f	=	"time-off" to "time-on" ratio
F	=	thrust, pounds
I	=	impulse, lb-sec
P	=	total mission time, seconds
r	=	control to disturbing torque ratio

Subscripts

0, 1, 2, 3...	=	conditions on points shown in various sketches
d or D	=	disturbing (except t_d, q. v. above)
c	=	control
Lc	=	limit cycle
u	=	undisturbed
co	=	cutoff locus position-axis intercept

Introduction

A spacecraft attitude control system must be capable of acquiring the correct orientation and maintaining it for long periods of time, and must have the capability of changing this orientation on command. The methods of actually accomplishing these tasks, the particular mechanisms used, the effects of various degrees of accuracy, and the techniques of simple system designs have received considerable attention in recent years. The types of control actuating mechanisms are numerous and can be classified into two large categories: momentum storage and momentum expulsion. The first class encompasses devices such as gyros and flywheels; the second encompasses the jet devices that are the topic of this paper. The systems analyzed are operating in an ON-OFF fashion, with a deadband, time delays, and external disturbing torques of varying magnitudes. Of particular interest are external disturbances of very low magnitudes. The behavior, convergence, stability, and energy consumption of the system are discussed as the external disturbances assume ever-increasing magnitudes from zero to relatively high values.

The Attitude Control Problem

The spacecraft attitude control problem originates from the need for definite and accurate orientation requirements. Because an exact orientation is not practical, acceptable position tolerances are assigned, and, as long as the spacecraft remains within this tolerance, termed deadband, no control actuation is required. The behavior of the system can be described in terms of phase-plane coordinates and its portraits plotted to determine its dynamical characteristics.

Zero External Torque

This case, having no external disturbing torques, is described by constant velocity trajectories in the deadband and constant acceleration trajectories outside the deadband. The operation of the system has been thoroughly covered in the open literature (1 to 8) and does not merit additional treatment here. It can be pointed out that the system requires control pulses given in terms of position and velocity according to the law

$$\pm\delta = k\,\dot{\theta} + \theta \qquad (1)$$

* Work presented here is based on a thesis undertaken in partial fulfillment of the requirements for the Degree of Engineer in Aerospace Engineering at the University of Southern California. The author wishes to acknowledge the services, cooperation, and advice of the Guidance Committee members, Professor B. Vernon, Chairman, Dr. Richard H. Edwards, and Dr. Raymond C. Binder. The author also expresses his gratitude to Mr. Perry L. Bailey for the programming of the equations.

otherwise it is unstable (2). The steady-state solution of this system is a limit cycle, termed undisturbed, shown in Fig. 1.

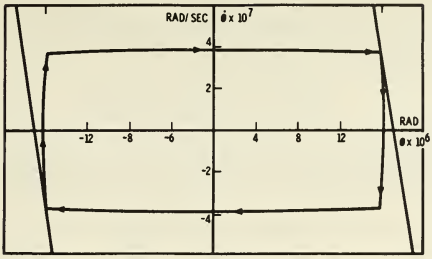


Figure 1. Undisturbed Limit Cycle ($\alpha_d = 0$)

Finite External Torque

This system is always under the influence of some torque and, hence, its motion does not follow any constant-velocity trajectories. The particular types of external torques that are assumed to exist here are those that result in constant unidirectional accelerations, normally experienced by spacecraft in interplanetary flights where the disturbances are caused by solar pressure, which implies that its motion is described by parabolic trajectories. In the deadband region, the system is subject to disturbing torque only, whereas the net torque outside this band is equal to the algebraic sum of the control and disturbing torques. The case where the relative magnitudes of the disturbances are such that the system does not traverse the entire deadband, thus requiring only one control pulse, has been adequately covered. (6, 7) The topic here is to investigate the behavior of a system with fixed control acceleration, but the disturbances vary in magnitude only. Of particular interest is the steady-state solution of the system if the disturbing acceleration is not sufficiently high to ensure that the system trajectories do not traverse the entire deadband. Before this is answered, the limit cycles analyzed are discussed briefly.

Limit Cycles

The particular limit cycles encountered in the jet attitude control systems are of the non-analytic type (11) with discontinuous slopes at the switching points; hence, they do not lend themselves to generalized analytical treatments. Numerical methods which proceed point-by-point can be used with the end points of each trajectory forming the initial conditions of the following trajectory. Limit cycles reached in systems of concern here are characterized by the operation of the jets at their minimum possible pulse widths. Because extended space missions will require long operations in the limit-cycle region, its energy expenditure must be minimized. For jet systems, this expenditure is measured in terms of impulse, which, in turn is translated into propellant weight, and is determined by the value of the jet thrust and its total operating time given by

$$I = F\, t_{on} \tag{2}$$

Since a mission is composed of control on and off times it becomes apparent that if the off-time is a maximum and the on-time a minimum, the energy consumption will be minimized. This is accomplished by a slight manipulation of Eq. (2).

$$I = F\, t_{on}\, \frac{t_{on} + t_{off}}{t_{on} + t_{off}} = F \cdot P \cdot \frac{1}{1+f} \tag{3}$$

If a quantity, termed energy ratio, is defined as

$$ER = \frac{1}{1+f} \tag{4}$$

the impulse equation becomes

$$I = F \cdot P \cdot (ER) \tag{5}$$

Therefore, the energy expenditure for a jet system with given thrust and mission duration will depend on the magnitude of the newly defined parameter, energy ratio.

The energy ratio of a disturbed limit cycle is uniquely defined by the ratio of the disturbing and the control accelerations (6), that is

$$\frac{\alpha_d}{\alpha_c} = \frac{1}{1+f} = ER \tag{6}$$

which implies a linear variation of ER with α_d, the slope given by $1/\alpha_c$.

The energy ratio of the undisturbed limit cycle depends on the switching function slope, the optimum value given by (6)

$$k_{opt} = \frac{t_o + \sqrt{t_o^2 + 32(\delta/\alpha_c)}}{4} \approx \sqrt{2\frac{\delta}{\alpha_c}} \tag{7}$$

The limit-cycle symmetry also has an effect on its energy ratio, as shown by use of the off-time to on-time ratio given by (6)

$$f = \frac{\alpha_c\, \theta}{\dot\theta_1\, \dot\theta_2} \tag{8}$$

where $\dot\theta_1$ is the velocity on the positive side of the cycle and $\dot\theta_2$ is the velocity on the negative side. Differentiating f with respect to $\dot\theta_1$, equating it with zero, and utilizing the relationship

$$\dot\theta_2 = \dot\theta_1 - \alpha_c\, t_{on} \tag{9}$$

$$\frac{df}{d\dot\theta_1} = -\frac{\alpha_c \theta}{\dot\theta_1\, \dot\theta_2}\left\{\frac{1}{\dot\theta_1} + \frac{1}{\dot\theta_2}\right\} = 0 \tag{10}$$

$$\dot\theta_1 = -\dot\theta_2$$

Therefore, the energy requirements are a maximum $[(d^2f)/(d\,\dot\theta_1^2) < 0]$ when the limit cycle is symmetric about the position axis.

Of special interest is to determine the relationships between the energy ratios of the disturbed

and undisturbed limit cycles. First, the point at which the two energy ratios are equal is determined.

From Eq. (6)

$$(ER)_D = \frac{\alpha_d}{\alpha_c} = \frac{1}{r} \tag{11}$$

from Eq. (A9) of Ref. 6

$$f = \frac{2(2k-t_r-t_d)}{t_d(2k-t_d)}\left\{\frac{2(\delta/\alpha_c)\ (2k-t_r-t_d)}{t_d(2k-t_d)} - (k-t_r)\right\} \tag{12}$$

which can be approximated by

$$f = \frac{2}{t_d}\left\{\frac{2\delta - kt_d\alpha_c}{\alpha_c t_d}\right\} \tag{13}$$

In any limit cycle operation, the off-time is much larger than the on-time, that is $f \gg 1$, which permits the approximation

$$(ER)_u \approx \frac{1}{f} \tag{14}$$

Therefore, the two energy ratios will be equal if the values of the disturbing and the control torques satisfy the conditions

$$r = f = \frac{4\delta - 2kt_d\alpha_c}{\alpha_c t_d^2} \tag{15}$$

The other point of interest is where control and disturbing accelerations are related in such a manner that the system traverses the entire deadband without intersecting the switching line at the left-hand side, (termed "optimum disturbed limit cycle" in Ref. 6). The ratio of control to disturbing accelerations here must satisfy the condition

$$r = \frac{4}{t_d}\sqrt{\frac{\delta}{\alpha_d}} \tag{16}$$

If the "on" and "off" times are related to the limit-cycle velocity $\dot{\theta}_{Lc}$ and the amplitude of oscillation θ_{Lc}, which is approximately equal for both types of limit cycles at this point, the undisturbed limit cycle energy ratio becomes

$$(ER)_u = 4\alpha_d/\alpha_c \tag{17}$$

which, when combined with Eq. (11), gives

$$\frac{(ER)_u}{(ER)_D} = \frac{4(\alpha_d/\alpha_c)}{(\alpha_d/\alpha_c)} = 4 \tag{18}$$

Thus, it is established that when the disturbing acceleration is very large, higher than that satisfying Eq. (15), the energy requirements of the disturbed limit cycle exceed those of the undisturbed. As the disturbing acceleration decreases, the energy requirements decrease linearly and become equal to one-fourth of the undisturbed limit cycle, which represents a zero disturbance. This critical point is reached when the ratio of the control to disturbing accelerations satisfies Eq. (16). The disturbing acceleration at that point is sufficient to reverse the system velocity just before it intersects the switching function on the left-hand side. The question here is the behavior of the system if the disturbing acceleration continues to decrease.

Ultralow Disturbance Region

This is the region bounded by zero and the critical value of the disturbing acceleration as defined by Eq. (16). Mathematically, it is given by

$$0 \le \alpha_d \le \frac{\alpha_c^2 t_d^2}{16\delta} \tag{19}$$

During system operation, the trajectories under the influence of the disturbing acceleration will intersect the left-hand side switching function. In the limit-cycle region the jets operate at their minimum possible pulse width, thus, when the trajectory intersects the switching function at the left-hand side of the deadband, the generated pulse width will be at, or near, its minimum value. The velocity of the system at the end of this pulse will be the algebraic sum of that generated by the control pulse, and the final velocity of the previous trajectory. After the control pulse is turned off, the system will have a positive velocity that will be further increased under the influence of the positive disturbing acceleration until it intersects the right-hand switching function, where another pulse is generated. The system now continues until it either converges into a stable limit cycle or, perhaps, diverges, in which case it is unstable.

Investigation of this special region commences with the derivation of the equations of motion given in Appendix A; Appendix B discusses the equations.

The phase plane is divided into three regions by the two switching functions (Fig. 14) on which the derivation of the equations of motion is based. In the deadband, the system is subjected to the influence of the disturbing acceleration only; outside the deadband, it is subjected to an acceleration that is the algebraic sum of the control and disturbing accelerations. This net acceleration is greater on the left side because the control and the disturbance act in the same direction, whereas they oppose each other on the right-hand side. Thus, a limit-cycle pulse generates a velocity increment of the left-hand side greater than that of the right-hand side because they both represent minimum-time trajectories.

Limit-Cycle Pulse Size

The possibility of three minimum pulse sizes exists because the disturbance may aid or oppose the control and there is also the case of zero disturbance. The pulse size is measured by the limit cycle velocity, which is evaluated as follows:

Zero Disturbing Torque--This is given by (1)

$$\dot{\theta}_{Lc} = \frac{\alpha_c t_d(2k - t_d)}{2(2k - t_r - t_d)} \tag{20}$$

Finite Disturbing Torque--The right-hand side pulse where control and disturbance act in opposing directions is given by [6, 7]

$$\dot{\theta}_{Lc} = \frac{(\alpha_c - \alpha_d)t_d(2k - t_d) + \alpha_d t_r(2k - t_r)}{2(2k - t_r - t_d)} \quad (21)$$

The left-hand side pulse is obtained by assuming that

$$\dot{\theta}_8 = -\dot{\theta}_6 = \dot{\theta}_{Lc} \quad (22)$$

which, when used in conjunction with Eqs. (A29) and (A32) gives, after a few minor manipulations,

$$2(k - t_d)\dot{\theta}_{Lc} + (\alpha_c + \alpha_d)t_d(t_d - 2k) + 2(\delta + \theta_6) = 0 \quad (23)$$

The term $(\delta + \theta_6)$ is evaluated by combining Eqs. (A21), (A26), and (A27), which results in

$$2(\delta + \theta_6) = (k - t_r)\dot{\theta}_{Lc} + \tfrac{1}{2}\alpha_d t_r(2k - t_r) \quad (24)$$

Substituting Eq. (24) into Eq. (23) and solving for $\dot{\theta}_{Lc}$, the limit-cycle velocity is given by

$$\dot{\theta}_{Lc} = \frac{(\alpha_c + \alpha_d)t_d(2k - t_d) - \alpha_d t_r(2k - t_r)}{2(2k - t_r - t_d)} \quad (25)$$

Equation (25) differs from Eq. (21) only by the sign of the disturbing acceleration and Eq. (20) can be obtained immediately from both Eq. (21) and Eq. (25) simply by letting the disturbance go to zero.

Discussion of System Motion

The system is of interest only if it is stable, which manifests itself by converging to a stable limit cycle. If a limit cycle exists, it will be characterized by control pulses on the left-hand side whose magnitude is defined by Eq. (25) and on the right-hand side by pulses of magnitude defined by Eq. (21).

Assume that the system begins with a velocity as defined by Eq. (25) on the left-hand side of the deadband, that is, a positive velocity corresponding to an exact symmetric limit-cycle pulse. As the system traverses the deadband to the right, it is accelerated by the disturbance until it intersects the right-hand side switching function. The right-hand side pulse generates a velocity increment defined by Eq. (21). Whether this velocity increment is of such a magnitude as to precisely reverse the velocity symmetrically (thus retracing the cycle on the left-hand side) can be shown by a close inspection of Eqs. (21) and (25), which define the magnitude of the right- and left-hand side limit-cycle pulses, respectively.

Recalling that, if the disturbing acceleration is large, the system will not reach the left-hand side of the deadband, and also realizing that this is an ultralow disturbance region permits the approximation

$$\alpha_c \gg \alpha_d \quad (26)$$

in which case, it is stated that

$$\alpha_c + \alpha_d \approx \alpha_c - \alpha_d \approx \alpha_c \quad (27)$$

An additional assumption made here (discussed at length in Ref. 6) is that start and cutoff time delays are approximately equal, i.e.,

$$t_r \approx t_d \quad (28)$$

When this assumption is incorporated in Eqs. (21) and (25), both become identical to

$$\dot{\theta}_{Lc} \approx \tfrac{1}{2}\alpha_c t_d \quad (29)$$

Equation (29) states that in the ultralow disturbance which is, of course, an approximation and must be treated as such. It indicates, however, that left- and right-hand control pulses are approximately equal in magnitude. The system motion is affected by the disturbance in the deadband, which will cause the velocity to increase from left to right (above the position axis) and decrease from right to left (below the position axis).

This indicates that the control pulse will not reverse the velocity symmetrically on the right-hand side if it is greater than that specified by Eq. (21) and approximated by Eq. (29); i.e., at the end of the pulse the velocity below the position axis will be lower in absolute value than that above the position axis. This mode of operation should continue, and as the system goes through each cycle it should not retrace its previous trajectories, but, instead, the deadband trajectories should be shifting upward with each trajectory passing a little higher in the positive velocity direction than the previous trajectories. This should continue, provided that the system traverses the entire deadband and gives control pulses on both sides.

In Appendix B, the conditions required for the trajectory to traverse the deadband from right to left against the disturbing accelerations are evaluated. The exact velocity at which this will occur can be calculated if the locus of the cutoff points on the lower right-hand side is considered in conjunction with Eq. (B6).

The cutoff points form a locus parallel to the switching function and defined by a vertical distance equal to the product of the cutoff delay t_d and the net acceleration. After applying the proper conditions for the right-hand side, this locus is given by

$$\delta = \tfrac{1}{2}t_d(\alpha_c - \alpha_d)(2k - t_d) + k\dot{\theta} + \theta \quad (30)$$

which, when combined with Eq. (B6), will define the magnitude of the velocity at the right-hand cutoff point, giving a trajectory that will not intersect the left-hand switching function. After a few simple and straight-forward algebraic manipulations, the following equation is obtained.

$$\dot{\theta}_4 = -\alpha_d k - \sqrt{4\alpha_d\delta - \alpha_d t_d(\alpha_c - \alpha_d)(2k - t_d)} \quad (31)$$

When this equation is satisfied the system will not give a control pulse on the left-hand side, and after it has reached the upper right-hand side, it will have a velocity lower than that of all the previous trajectories. After a right-hand pulse is given, the velocity on the lower side will be higher (in absolute value) than in the previous cases. This will cause the system to require pulses on the left-hand side again. The procedure from here should be a repetition of the upward motion of the trajectories across the deadband until the conditions of Eq. (31) are met again. This, then, should be the periodic solution of the system.

Therefore, the periodic solutions of the system in this ultralow disturbance region must consist of limit cycles composed of many cycles with two pulses, one for each side, and one cycle with a single pulse on the right-hand side only; similar to a fully disturbed limit cycle. Futhermore, the number of the double-pulse cycles must decrease with increasing disturbances because this upward shifting of the cycles is caused by the velocity change across the deadband, which increases with the disturbance magnitude. That this is the case was verified by use of a digital computer program formulated to aid in the computation. The results obtained from numerical examples fully verify the above discussion.

Numerical Results--System Convergence

The behavior of the system in the ultralow disturbance region discussed previously was verified by the use of a digital computer program. A typical set of input data was assumed and other information about the system was calculated; all the pertinent numerical parameters are given in Table 1.

GIVEN PARAMETERS	
Spacecraft moment of inertia, slug-ft^2	10,000
Control moment arm, ft	5
Deadband magnitude, rad	0.17×10^{-4}
Control jet thrust, lb	0.1
Time delays, sec	
Start	0.010
Cutoff	0.015
CALCULATED PARAMETERS	
Control torque, ft-lb	0.5
Control acceleration, rad/sec^2	0.5×10^{-5}
Optimum switching function slope, sec	0.83
Critical disturbing acceleration, rad/sec^2	0.222×10^{-8}

TABLE 1 SAMPLE NUMERICAL EXAMPLE

First, a few points were checked with zero and high ($\alpha_d > 0.22 \times 10^{-8}$ rad/sec^2) disturbance. The corresponding limit cycles were established (Figs. 1 and 2). In the ultralow disturbance region, the disturbing acceleration was varied

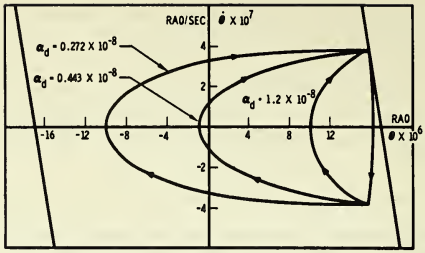


Figure 2. Disturbed Limit Cycles ($\alpha_d > 0.222 \times 10^{-8}$ Rad/Sec^2)

over a wide range. The obtained results (shown in Table 2), fully verified the predictions made in the previous sections. Phase portraits of the limit cycles are shown in Figs. 3 through 10.

The results verify the prediction that in this ultralow disturbance region the system is stable, and its steady-state periodic solution is represented by limit cycles that are composed of multiple loops; i.e., they are composed of a certain number of double-pulse cycles and always one single-pulse cycle. Thus, in the cases tested, the number of double-pulse cycles goes from 1 to 16, as the disturbing acceleration goes from 0.20×10^{-8} rad/sec^2 to 0.01×10^{-8} rad/sec^2, whereas the number of single-pulse cycles remains 1 over the entire region.

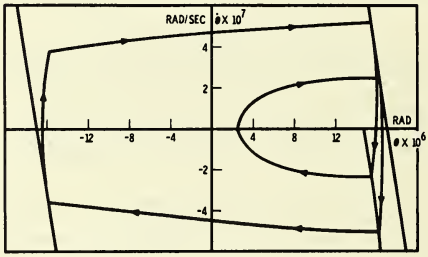


Figure 3. Limit Cycle For $\alpha_d = 0.2 \times 10^{-8}$ Rad/Sec^2

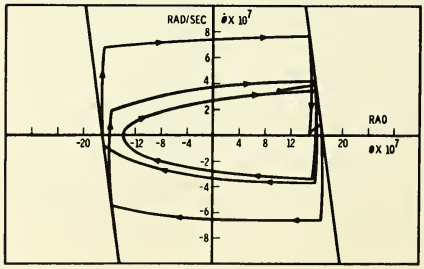


Figure 4. Transient Response--First Three Cycles: $\alpha_d = 0.2 \times 10^{-8}$ Rad/Sec^2

Disturbing Acceleration rad/sec^2 x 10^{-8}	0.20	0.18	0.13	0.12	0.105	0.08	0.05	0.01
Number of double pulse cycles	1	1	1	2	3	3	5	16
Number of single pulse cycles	1	1	1	1	1	1	1	1
Time on, sec	0.04507	0.04545	0.04507	0.0751	0.1052	0.1052	0.1653	0.473
Time off, sec	378	414	575	629	729	931	1,518	4,166
Energy ratio x 10^{-3}	0.1192	0.1097	0.0785	0.119	0.144	0.113	0.109	0.1134
Normalized ER	0.67	0.615	0.44	0.67	0.81	0.635	0.613	0.638

Undisturbed limit cycle: $\alpha_d = 0$; ER = 0.178×10^{-3}; normalized: 1.0

Optimum disturbed limit cycle: $\alpha_d = 0.222 \times 10^{-8}$; ER = 0.0445×10^{-3}; normalized: 0.25

TABLE 2 ULTRALOW DISTURBANCE REGION: LIMIT CYCLE CHARACTERISTICS WITH VARIOUS LEVELS OF DISTURBING ACCELERATIONS

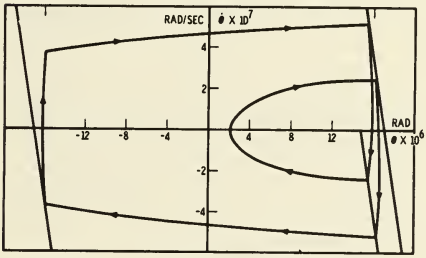


Figure 5. Steady-State Phase Portrait ($\alpha_d = 0.2 \times 10^{-8}$ Rad/Sec^2)

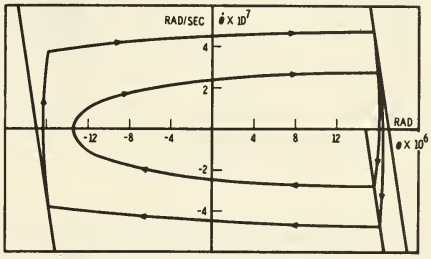


Figure 7. Limit Cycle For $\alpha_d = 0.13 \times 10^{-8}$ Rad/Sec^2

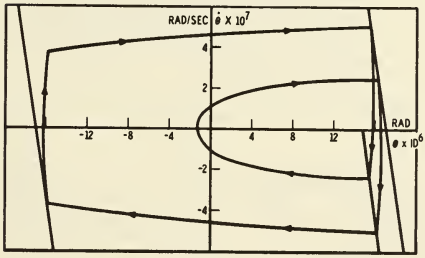


Figure 6. Limit Cycle For $\alpha_d = 0.18 \times 10^{-8}$ Rad/Sec^2

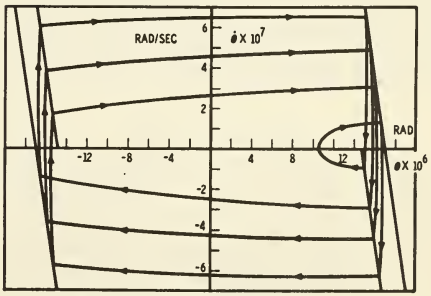


Figure 8. Limit Cycle For $\alpha_d = 0.10 \times 10^{-8}$ Rad/Sec^2

The answer, therefore, to the question posed originally concerning what will happen if the system intersects the left-hand switching function, is that the system will settle into a new limit cycle that is composed of two cycles, one double-pulse cycle and one single-pulse cycle, as shown in Fig. 3. The stability of this cycle was tested by starting it at some arbitrary point; Fig. 4 shows the first three cycles of the transient response, but, some time later, cycles 40 to 50 were checked and plotted in Fig. 5. The result is identical to that shown in Fig. 3.

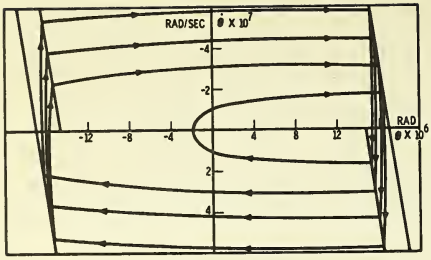


Figure 9. Limit Cycle For $\alpha_d = 0.08 \times 10^{-8}$ Rad/Sec^2

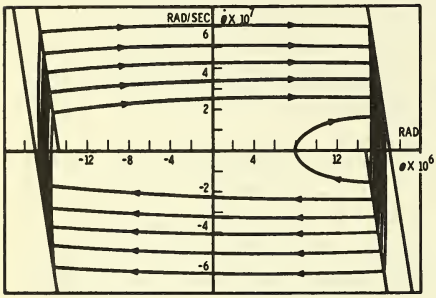


Figure 10. Limit Cycle For $\alpha_d = 0.05 \times 10^{-8}$ Rad/Sec^2

As the disturbance continues to decrease, the point of intersection of the single-pulse cycle with the position axis moves to the left, as shown in Figs. 5, 6, and 7, and when it intersects the left-hand switching function again, it changes into a limit cycle with two double-pulse cycles and one single-pulse cycle. As the disturbance continues to decrease, this process is repeated and every time the inner single-pulse cycle intersects the left-hand switching function, an additional double-pulse cycle is generated. This is shown pictorially in the phase portrait plots in Figs. 3 through 10 and numerically in Table 2.

Energy Requirements

The energy requirements, being a measure on the "on" to "off" time ratio, are measured by the energy ratio defined by Eq. (4). As indicated previously and as verified by the numerical results shown in Table 2, the "on" time of each pulse is almost a constant; this is evident, from Table 2 for the cases that are composed of the same number of double-pulse cycles, e. g., the first three cases. This is not the case, however, with the "off" time, which increases with decreasing disturbance until the limit cycle changes discontinuously into one with more cycles. Therefore, as the disturbing acceleration decreases, the energy ratio will decrease until the single-pulse cycle intersects the left-hand switching function, at which point the energy ratio increases discontinuously. This process will continue, the energy ratio decreasing with decreasing disturbance, and when the pattern of the limit cycle is altered, the energy ratio increases discontinuously again. This is verified from the numerical results given in Table 2, which also show that although the energy ratio goes through these variations, it never exceeds the value associated with the completely undisturbed limit cycle case.

Conclusions

The following main conclusions can be reached about the behavior of the system in the ultralow disturbance region:

1. The system has stable periodic solutions that are represented by multiple-cycled limit cycles. These are composed of one single-pulse, or inner cycle, and several double-pulse, or outer cycles, whose number increases with decreasing disturbance.
2. The limit-cycle energy ratio, although displaying a discontinuous and somewhat erratic behavior, remains always below that of the completely undisturbed case.
3. The limit cycle corresponding to a zero disturbance can be considered as being composed of an infinite number of double-pulse cycles and zero single-pulse cycles. Because these double-pulse cycles coincide with one another and are symmetric, they have the maximum energy ratio; hence, the energy ratio of the undisturbed limit cycle represents an upper bound for the system energy expenditure in the ultralow disturbance region.
4. When a system is designed and provided with an energy capacity as specified by a zero disturbance, then any increase in the disturbance magnitude will result in energy savings. These savings will continue until the disturbing acceleration exceeds the value specified by Eq. (15), beyond which the disturbances cause an excessive consumption of energy as shown in Fig. 11.

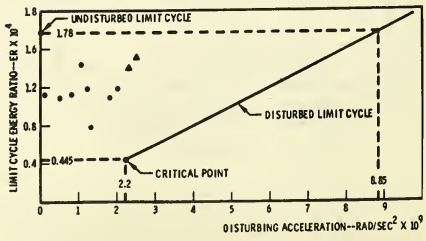


Figure 11. Limit Cycle Energy Ratio Variation With Disturbances

A word of caution must be given here about the possibility of nonunique limit cycles in some regions. When the disturbing acceleration was decreased below the critical value required for a fully disturbed limit cycle operation, it immediately resulted in a multiple-loop cycle. If the procedure is reversed, however, the change does not take place at the same point, but a hysteresis-type effect is observed. In the system used here, the possibility of multiple solutions is indicated by Figs. 12 and 13, which should be in the fully disturbed region. This can be interpreted as a peculiar region where both "soft" and "hard" excitations are possible. The energy requirements of the hard cycles are higher than those of the soft, which make them undesirable. They can be eliminated from a system such as this, if provisions are made for erasing the left-hand switching function; under such conditions the system will always settle in a unique limit cycle, although it may undergo a few overshoots. Before the left-hand switching function can be eliminated, however, the disturbing acceleration must be greater than the critical value specified by Eq. (16), which forms the boundary of the ultralow disturbance region.

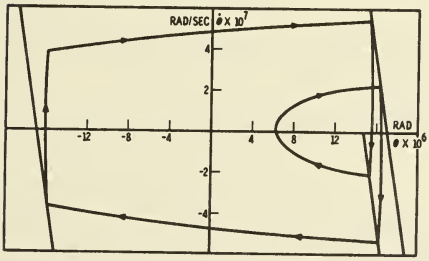


Figure 12. "Hard" Limit Cycle For $\alpha_d = 0.23 \times 10^{-8}$ Rad/Sec^2

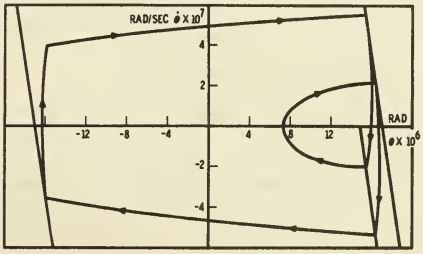


Figure 13. "Hard" Limit Cycle For $\alpha_d = 0.25 \times 10^{-8}$ Rad/Sec^2

Appendix A:
Derivation of the Equations of Motion

The problem begins with a set of given initial conditions as described by the position and velocity coordinates at time zero, θ_o and $\dot{\theta}_o$. Other given parameters are the control parameters, α_c, α_d, δ, k, t_r, t_d; the procedure is a point-by-point step beginning with 0, as given by (θ_o, $\dot{\theta}_o$), going through points 1, 2, . . . , 10, as shown in Fig. 14.

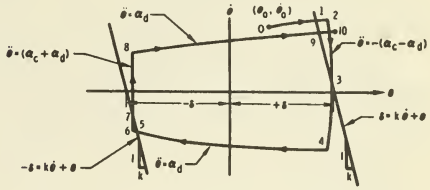


Figure 14. Transient Motion (Converging) Phase Portrait

The equations of motion are: Path 0 - 1: $\ddot{\theta} = \alpha_d$. At point 1, the trajectory will intersect the switching function; the state of the system is then determined by the simultaneous solution of the following equations:

$$\delta = k\dot{\theta}_1 + \theta_1 \tag{A1}$$

$$\theta_1 = \theta_o + \frac{\dot{\theta}_1^2 - \dot{\theta}_o^2}{2\alpha_d} \tag{A2}$$

Substituting the value of θ_1, from (A1) into (A2), gives, after a minor transformation,

$$(\delta - k\dot{\theta}_1 - \theta_o)\, 2\alpha_d = \dot{\theta}_1^2 - \dot{\theta}_o^2 \tag{A3}$$

$$\dot{\theta}_1^2 + 2\alpha_d k\dot{\theta}_1 - 2\alpha_d (\delta - \theta_o) - \dot{\theta}_o^2 = 0 \tag{A4}$$

The conditions at point 1 are

$$\dot{\theta}_1 = -\alpha_d k + \sqrt{\alpha_d^2 k^2 + 2\alpha_d (\delta - \theta_o) + \dot{\theta}_o^2} \tag{A5}$$

$$\dot{\theta}_1 = \delta - k\dot{\theta}_1 \tag{A6}$$

$$t_{o-1} = \frac{|\dot{\theta}_1 - \dot{\theta}_o|}{\alpha_d} \tag{A7}$$

The conditions at point 2 are constrained by the given time delay t_r; i.e., at point 1 an on signal is given, but control actuation does not materialize until t_r seconds have elapsed. Therefore, the conditions at point 2 are

$$\dot{\theta}_2 = \dot{\theta}_1 + \alpha_d t_r \tag{A8}$$

$$\theta_2 = \theta_1 + \frac{\left(\dot{\theta}_2^2 - \dot{\theta}_1^2\right)}{2\alpha_d} \tag{A9}$$

$$t_{1-2} = t_r \tag{A10}$$

The acceleration for the 2-4 trajectory is

$$\ddot{\theta} = -(\alpha_c - \alpha_d) \tag{A11}$$

The conditions of point 3 are determined by the intersection of the switching function and the new constant acceleration trajectory is

$$\delta = k\dot{\theta}_3 + \theta_3) \tag{A12}$$

$$\theta_3 = \theta_2 - \frac{\left(\dot{\theta}_3^2 - \dot{\theta}_2^2\right)}{2(\alpha_c - \alpha_d)} \tag{A13}$$

Following the procedure set forth in point 1, i.e., substituting (A12) into (A13) and transforming, gives

$$\dot{\theta}_3^2 - 2k(\alpha_c - \alpha_d)\,\dot{\theta}_3 + 2(\alpha_c - \alpha_d)(\delta - \theta_2) - \dot{\theta}_2^2 = 0 \tag{A14}$$

which results in

$$\dot{\theta}_3 = k(\alpha_c - \alpha_d) - \sqrt{k^2(\alpha_c - \alpha_d)^2 - 2(\alpha_c - \alpha_d)(\delta - \theta_2) + \dot{\theta}_2^2} \tag{A15}$$

The negative square root being accepted as the valid solution, because the possibility of $\dot{\theta}_3 < 0$ must be allowed for [this is assured by Eq. (A15)], the other conditions for point 3 are

$$\theta_3 = \delta - k\dot{\theta}_3 \tag{A16}$$

$$t_{2-3} = \frac{|\dot{\theta}_3 - \dot{\theta}_2|}{\alpha_c - \alpha_d} \tag{A17}$$

The conditions at point 4 are constrained by the given time delay t_d; i.e., at point 3 an off signal is given, but this does not take effect until t_d seconds have elapsed. Therefore, the conditions at point 4 are

$$\dot{\theta}_4 = \dot{\theta}_3 - (\alpha_c - \alpha_d)\,t_d \tag{A18}$$

$$\theta_4 = \theta_3 - \frac{\left(\dot{\theta}_4^2 - \dot{\theta}_3^2\right)}{2(\alpha_c - \alpha_d)} \tag{A19}$$

$$t_{3-4} = t_d \tag{A20}$$

As indicated in Fig. 14, the system now proceeds to point 5 at the other end of the deadband. Before this happens, i.e., for trajectory 4-5 to be possible, it is necessary for $\dot{\theta}_4$ to be negative.

If $\dot{\theta}_4 > 0$, then the procedure from point zero is repeated as many times as necessary, as shown in Fig. 15, until $\dot{\theta}_4 < 0$. This is defined as the "reticence" phenomenon and whether or not it will occur depends on a variety of parameters, but primarily on the values of $\dot{\theta}_2$, α_c, α_d, k, and t_d.

The system will proceed to point 5, as shown in Fig. 15, where the conditions are described by the intersection of the constant acceleration ($\ddot{\theta} = \alpha_d$) trajectory and the switching function, as shown by the following equations

$$-\delta = k\dot{\theta}_5 + \theta_5 \tag{A21}$$

$$\theta_5 = \theta_4 + \frac{\left(\dot{\theta}_5^2 - \dot{\theta}_4^2\right)}{2\alpha_d} \tag{A22}$$

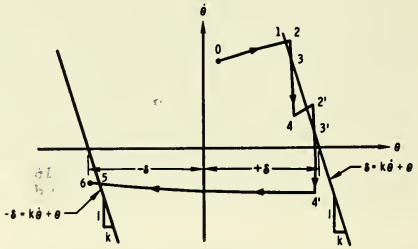


Figure 15. Reticence Phenomenon

which, when solved simultaneously for $\dot{\theta}_5$, give

$$\dot{\theta}_5 = -k\alpha_d - \sqrt{k^2\alpha_d^2 - 2\alpha_d(\delta + \theta_4) + \dot{\theta}_4^2} \tag{A23}$$

Here again the negative square root represents the only acceptable solution, for $\dot{\theta}_5$ must always be negative and Eq. (A23) assures it. The position coordinate and the time for point 5 are

$$\theta_5 = -\delta - k\dot{\theta}_5 \tag{A24}$$

$$t_{4-5} = \frac{|\dot{\theta}_5 - \dot{\theta}_4|}{\alpha_d} \tag{A25}$$

The conditions of point 6 are related to those of point 5 in a manner similar to that existing between points 1 and 2; i.e., a constant time (t_r) constraint.

Therefore, point 6 is described by

$$\dot{\theta}_6 = \dot{\theta}_5 + \alpha_d\,t_r \tag{A26}$$

$$\theta_6 = \theta_5 + \frac{\left(\dot{\theta}_6^2 - \dot{\theta}_5^2\right)}{2\alpha_d} \tag{A27}$$

$$t_{5-6} = t_r \tag{A28}$$

Continuing with this point-by-point procedure and allowing for the "reticence" phenomenon to occur between points 5 and 8 unless $\dot{\theta}_8 > 0$, the following relations are derived.

Point 7: $\dot{\theta}_7 = -k(\alpha_c + \alpha_d)$

$$+\sqrt{k^2(\alpha_c+\alpha_d)^2 - 2(\alpha_c+\alpha_d)(\delta+\theta_6)+\dot{\theta}_6^2} \quad \text{(A29)}$$

$$\theta_7 = -\delta - k\dot{\theta}_7 \quad \text{(A30)}$$

$$t_{6-7} = \frac{|\dot{\theta}_7 - \dot{\theta}_6|}{\alpha_c + \alpha_d} \quad \text{(A31)}$$

Point 8: $\dot{\theta}_8 = \dot{\theta}_7 + (\alpha_c + \alpha_d)\, t_d$ (A32)

$$\dot{\theta}_8 = \theta_7 + \frac{\left(\dot{\theta}_8^2 - \dot{\theta}_7^2\right)}{2(\alpha_c + \alpha_d)} \quad \text{(A33)}$$

$$t_{7-8} = t_d \quad \text{(A34)}$$

If $\dot{\theta}_8 < 0$, then the steps between points 5 and 8 are repeated until $\dot{\theta}_8 > 0$, otherwise the trajectory 8-9-10 is not possible because it requires a positive velocity.

Point 9: $\dot{\theta}_9 = -k\alpha_d$

$$+\sqrt{k^2\alpha_d^2 + 2\alpha_d(\delta - \theta_8) + \dot{\theta}_8^2} \quad \text{(A35)}$$

$$\dot{\theta}_9 = \delta - k\dot{\theta}_9 \quad \text{(A36)}$$

$$t_{8-9} = \frac{|\dot{\theta}_9 - \dot{\theta}_8|}{\alpha_d} \quad \text{(A37)}$$

Point 10: $\dot{\theta}_{10} = \dot{\theta}_9 + \alpha_d\, t_r$ (A38)

$$\dot{\theta}_{10} = \theta_9 + \frac{\left(\dot{\theta}_{10}^2 - \dot{\theta}_9^2\right)}{2\alpha_d} \quad \text{(A39)}$$

$$t_{9-10} = t_r \quad \text{(A40)}$$

Appendix B
Discussion of the Equations of Motion

The convention followed in this analysis is that positive accelerations result in trajectories concave to the right, whereas negative accelerations are represented by trajectories concave to the left. This is a working basis and is in agreement with physical realities where a positive acceleration will tend to increase a positive velocity, and conversely. This convention has rendered the external disturbing acceleration (α_d) positive, which will increase all positive and will decrease all negative velocities. Therefore, for the upper portion of the phase plane shown in Fig. 14, as the trajectories go from left to right the velocities increase, whereas in the negative velocity region, the lower portion, the velocities decrease as the trajectories go from right to left, that is,

$$\dot{\theta}_2 > \dot{\theta}_1 > \dot{\theta}_0 \quad \text{(B1)}$$

$$|\dot{\theta}_4| > |\dot{\theta}_5| > |\dot{\theta}_6| \quad \text{(B2)}$$

This causes the disturbing acceleration to have a damping effect in the negative velocity region, whereas it has an exciting effect in the positive velocity region. Therefore, the intersection of all constant acceleration trajectories and the switching function on the right-hand side is ensured. This cannot be ensured, however, for the switching function on the left side, because the disturbing acceleration acting in a direction such as to reduce the velocity, may reduce it to zero and reverse the direction of the motion before the required distance has been traversed. This is shown schematically in Fig. 16, where the trajectories originating at point 4 under different acceleration conditions are indicated. In effect, it is this point 4 which may have a velocity and position such as to prevent the trajectory from reaching point 5. In Fig. 16, the trajectories followed under five different values of acceleration are indicated.

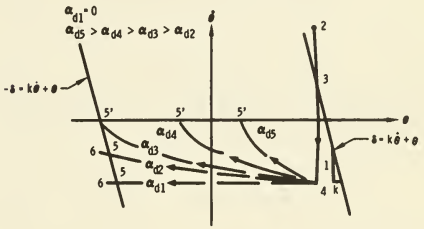


Figure 16. Effect of α_d on Trajectory Originating From Point 4

When the value of the external disturbing acceleration is that described schematically by $\alpha_d > \alpha_{d3}$ in Fig. 16, simultaneous solution of Eqs. (A21) and (A22) cannot be used for calculating the conditions at point 5, because the trajectory dictated by this acceleration will not intersect the switching function at the left. A similar situation can also exist if the value of the velocity at point 4 is below some critical value.

Whether the trajectory originating at point 4 will intersect the switching function will be determined by the values of $\dot{\theta}_4$, θ_4, δ, α_d, and k. Mathematically, the case in which these two curves do not intersect manifests itself in complex values for $\dot{\theta}_5$ as given by Eq. (A23). Therefore, when the solution of this equation indicates nonintersection by giving complex values to $\dot{\theta}_5$, then the equations for this path are altered slightly to properly reflect the new situation. This is handled by calculating the point of zero velocity furthest to the left, denoted by 5' in Fig. 16. Actually, points 5 through 8, as shown in Fig. 14, do not exist and a new

point 5' takes their place. The applicable equations are

$$\dot{\theta}_{5'} = \dot{\theta}_5 = \dot{\theta}_6 = \dot{\theta}_7 = \dot{\theta}_8 = 0 \qquad \text{(B3)}$$

$$\theta_{5'} = \theta_5 = \theta_6 = \theta_7 = \theta_8 \qquad \text{(B4)}$$

$$\theta_{5'} = \theta_4 - \frac{\dot{\theta}_4^2}{2\alpha_d} \qquad \text{(B5)}$$

The system proceeds to points 9, and 10 as shown in Fig. 14 and as represented by Eqs. (A35) through (A40), with the values of $\dot{\theta}_8$ and θ_8 in Eqs. (A35), (A36), and (A37) being replaced by those indicated by Eqs. (B3) and (B5).

Conditions Under Which Point 5' Exists

It was stated that nonintersection of the trajectory originating at point 4 of Fig. 16 with the switching function at the left manifests itself in complex values of the velocity at point 5, as evaluated by Eq. (A23). Therefore, for intersection, it is required that $\dot{\theta}_5$ be real, i.e.,

$$\alpha_d^2 k^2 + \dot{\theta}_4^2 - 2\alpha_d(\delta + \theta_4) \geq 0 \qquad \text{(B6)}$$

or

$$\alpha_d^2 k^2 + \dot{\theta}_4^2 \geq 2\alpha_d(\delta + \theta_4) \qquad \text{(B7)}$$

If the values of k, $\dot{\theta}_4$, θ_4 and are known, then the required value of the disturbing acceleration α_d is obtained by solution of Eq. (B6)

$$\alpha_d^2 - \frac{2}{k^2}(\delta + \theta_4)\alpha_d + \frac{\dot{\theta}_4^2}{k^2} \geq 0 \qquad \text{(B8)}$$

From which it is concluded that if α_d meets the conditions specified by

$$\alpha_d < \frac{\delta + \theta_4}{k^2} - \sqrt{\left[\frac{\delta + \theta_4}{k^2}\right]^2 - \frac{\dot{\theta}_4^2}{k^2}} \qquad \text{(B9)}$$

the trajectory originating at point 4 will intersect the switching function on the left given by $-\delta = k\dot{\theta} + \theta$, and thus point 5' will not exist.

References

1. Tavlor, W. L., Jr. and J. W. Smith, "An analytical approach to the design of an automatic discontinuous control system," NASA TN D-630 (April 1961).

2. Pistiner, J. S., "On-off control system for attitude stabilization of space vehicles," ARS J., Vol. 29, No. 4, pp 283-289 (April 1959).

3. Gaylord, R. S., and W. N. Keller, "Attitude control system using logically controlled pulses," Progress in Astronautics and Rocketry: Guidance and Control, edited by R. Roberson and J. S. Farrior (Academic Press, New York, 1962), Vol.8, pp 629-648.

4. Dahl, P. R., L. K. Herman, and G. T. Aldrich, "Limit cycles in reaction jet attitude control systems subject to external torque," Progress in Astronautics and Rocketry: Guidance and Control, edited by R. Roberson and J. S. Farrior (Academic Press, New York, 1962), Vol. 8, pp 599-627.

5. Nicklas, J. C., and H. C. Vivian, "Derived-rate increment stabilization: its application to the attitude control problem," TR 32-69, Jet Propulsion Laboratory, Pasadena, California (31 July 1961).

6. Haloulakos, V. E., "Analytic determination of thrust level and impulse requirements for spacecraft on-off jet attitude control systems," AIAA Preprint No. 63-237, presented at the AIAA Summer meeting, Los Angeles, California (June 17-20, 1963).

7. Haloulakos, V. E., "Thrust and impulse requirements for jet-attitude control systems," J. of Spacecraft and Rockets, Vol. 1, No. 1, pp 84-90 (January 1964).

8. Lange, B., "The drag-free satellite," AIAA J., Vol. 2, No. 9, pp 1590-1606 (September 1964).

9. Ku, Y. H., Analysis and Control of Nonlinear Systems (The Rolland Press, New York, 1958).

10. Davis, H. T., "The phase plane and its phenomena," Introduction to Nonlinear Differential and Integral Equations, (U.S. Atomic Energy Commission, September 1960), pp 267-308.

11. Minorsky, N., Nonlinear Oscillations, (Van Nostrand, New York, 1962.

No. 67-538

DESCRIPTION OF AN INERTIAL TEST FACILITY LOCATED 1,100 FEET BELOW THE SURFACE OF THE EARTH

by

W. M. WEIL, W. E. SRODE,
and
R. M. BURROWS
LTV Aerospace Corporation
Warren, Michigan

AIAA Paper
No. 67-538

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

10.05, 10.10

DESCRIPTION OF AN INERTIAL TEST FACILITY LOCATED 1, 100 FEET BELOW THE SURFACE OF THE EARTH

W. M. Weil, W. E. Srode, and R. M. Burrows
Missiles and Space Division-Michigan
LTV Aerospace Corporation
Warren, Michigan

Abstract

A unique facility for testing advanced inertial guidance and navigation systems and components has been developed more than 1, 100 feet below the earth's surface in an unused portion of the International Salt Company mines under Detroit. The laboratory site, carved out of natural rock salt, is at least 2 miles removed from the main shaft and active mining operations. Noteworthy advantages of the underground location are: 1) bedrock floor, providing extreme resistance to seismic disturbances, very low test pier construction costs, and no restriction on equipment size and weight; 2) constant 57° + 1°F ambient temperature, reducing heating and air-conditioning costs; 3) elimination of radio frequency interference; and 4) low laboratory construction costs. Because of its seismic stability, this facility is expected to provide the best natural test pad stability in the nation. Initial measurements show ambient vibration levels averaging from 1 to 5 micro-g's for frequencies below 20 cps and from 10 to 30 micro-g's in the 20-to-200 -cps range. Floor stability is excellent: A 10-minute tremor recorded during initial observations moved the bedrock floor only + 3 arc seconds; heavy traffic within 20 feet and personnel movement within 6 feet produced no measurable disturbances.

Introduction

The Missiles and Space Division of LTV Aerospace Corporation recently conducted a program at its Michigan facility to establish a capability to accurately test advance and state-of-the-art inertial guidance systems and components. This paper details the investigations, design considerations, and decisions that resulted in establishing a unique test facility inside an unused portion of an operating salt mine 1, 100 feet below the earth's surface (Figure 1). This test facility will undoubtedly contribute to the arts of inertial component testing; provide reference material for engineers and scientists involved in projects pertaining to the extreme accuracy of advanced inertial guidance or navigation systems; and add to current knowledge of cultural noise levels within the earth's crust. As presently planned, this facility will possess the nation's best natural seismic stability for testing.

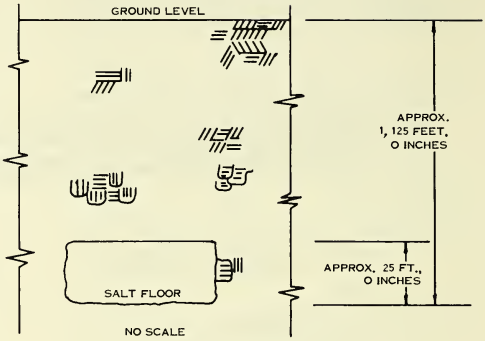


Figure 1. Site Depth

The science of seismic floor design and construction has been hard-pressed to keep pace with the rapid progress in performance capabilities of inertial sensors and systems. MSD-Michigan faced this problem in the early spring of 1966. At the time, all of our inertial device testing was performed on a simple test pad with minimal isolation from the surrounding environment. Vibration measurements indicated noise levels far in excess of the equipment threshold levels being specified for new programs at that time.

Ideally, the magnitudes of angular rotations of an inertial test pad and vibrations transmitted through the pad should be a lesser order than that of the smallest parameter being measured so as not to obscure the performance characteristics of the devices being tested.

Realizing that this goal might be impossible to reach, we arbitrarily established the minimum requirement that the test pad exhibit an angular tilt and ambient-vibration noise level not exceeding one-half of the magnitude of the smallest component parameters to be measured. This criterion established our test pad stability design goals as:

1. Angular stability such that the angular rotation of the pad about any of 3 mutually perpendicular axes shall not exceed 2 arc-seconds over a

24-hour period, and shall not exceed 5 arc-seconds during any 30-day period.

2. Ambient vibration noise level of no more than 10 micro-g's peak for all frequencies above 2 cps.

Four approaches toward obtaining the required test pad stabilities were considered as follows:

1. Building a test pad in an area having a seismic environment that would inherently meet the requirements.

2. Building a test pad using tilt and vibration isolation techniques.

3. Using precision measuring and recording instruments to monitor existing environments during test operations, then compensating the test data to eliminate the induced environmental effects.

4. Further searching for an existing test facility capable of meeting our requirements.

We began a study to determine whether there was any location within a 30-mile radius of the plant that was both remote from cultural noise and relatively close to solid rock. We found that the closest bedrock structure under the Detroit-Warren area resembles a huge bowl, the deepest portion being approximately 200 feet immediately below Detroit, gradually sloping upward but not approaching the surface within 40 to 50 miles of MAMP. This structure, coupled with the heavy manufacturing in the Detroit area, all but precluded our finding an inherently benign seismological environment in the immediate area.

During a brainstorming session, the feasibility and practicality of actually locating the inertial test pad and laboratory facilities in the International Salt Company mines under Detroit was discussed. We contacted the mine officials and, after being convinced that we were indeed serious, they agreed to allow us to visit the Salt Mine on an inspection trip. We expected an unusual trip, interesting, but not one to solve our problem. As question after question was answered satisfactorily, it became apparent that it would be desirable to lease space in an unused portion of the mine as a test pad site.

The salt was discovered in 1895, the salt mine (Figure 2)has been operating since 1910. The mine shaft and operations are located in southwestern Detroit and extend radially from the shaft southwest and west for about 2 miles. Mining operations have improved over the years and today salt mining is a well-controlled operation: New salt faces are prepared and blasted at night; the next morning the salt is hauled away. Since these blasts occur at the same time every day, testing is permitted without unknown interference from the blasts. This particular feature--along with the inherent environments such as constant temperature and low seismic vibration and noise--have made the mine a desirable site for conducting special projects and studies. Other scientific organizations have successfully conducted projects ranging from calibrating satellites to measuring the amount of the sun's cosmic rays that penetrate the earth at the depth of the mine. Because the mining operation uses the room and pillar system, the layout of the mine resembles a honeycomb: huge rooms--50 feet wide and 23 feet high--are carved out of the salt. Roof-supporting pillars between rooms are about the same dimensions, making this a highly stable structure. All these features made the mine an ideal location for constructing a test pad with the capabilities required for testing advanced inertial systems.

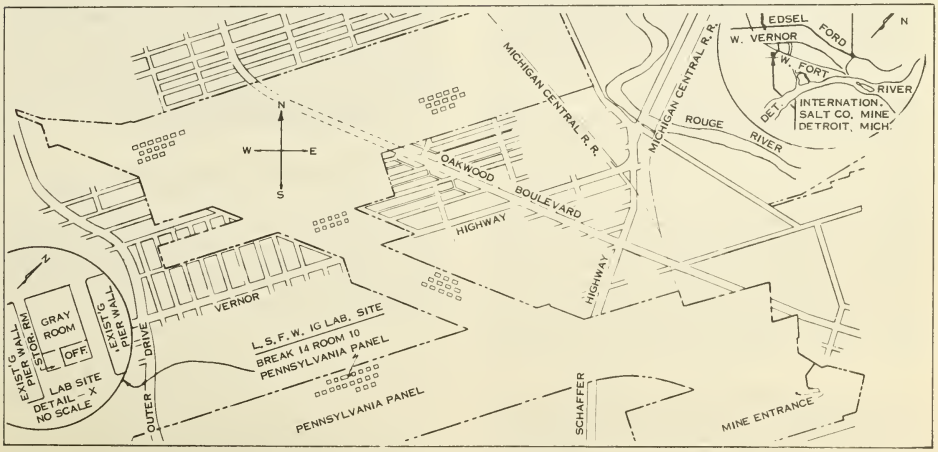


Figure 2. Location Map

67-538

The laboratory or test pad site selected is located about 5 minutes by jeep from the main shaft of the mine in a room that had not been used for 15 years. Both the 1-minute ride down the shaft to the mine -- 1, 100 feet below the surface -- and the 5-minute jeep ride from the main shaft are interesting experiences. Machine shops, maintenance shops, and huge pieces of mining equipment are along the route. Vehicle traffic on the "salt roads" is similar to surface traffic. Jeep-size vehicles and equipment can be lowered down the shaft fully assembled. Larger equipment is disassembled and lowered piece by piece.

Preliminary Evaluation of the Site

One visit convinced us of the potential suitability of the salt mine site for inertial testing. We then formulated a test program to quantitatively determine its environmental qualifications, including:

1. Ambient vibration level.

2. Angular stability of the mine floor.

3. Feasibility of establishing a north-south reference line using gyrocompass techniques (Polaris not being available as a primary north reference).

4. Ambient temperature.

5. Relative humidity.

6. Gaseous contaminant content, corrosive properties, and dust and salt particle count of the mine air.

Northrup Nortronics was engaged to make the first three determinations, under the direction of Dr. S. Okubo. Temperature and humidity measurements and air analyses were made by MSD personnel. The test program, conducted in July and August of 1966, proved that the site either met or exceeded our requirements.

Background Seismic Vibration

Simultaneous seismometer readings were taken on the floor about 6 inches above bedrock in each of three orthogonal axes: vertical, horizontal north-south, and horizontal east-west. These data were collected using Hall-Sears Model HS-10 seismometers having a coil resistance of 210, 000 ohms and a spring-suspended mass with a natural frequency of 2 Hertz. Each coil is wound around an inertial mass which is spring-suspended around a magnet attached to the case; load resistance across the coil provides damping. The output voltage or seismic reading is therefore proportional to the relative velocity of the coil and magnet.

These seismometers were calibrated on University of Michigan's Geophysics Laboratory linear air spring suspension shake table: an electromagnetic transducer, analogous to a moving coil loudspeaker, which contains a dc field coil to supply magnetic flux, a force coil that moves the shake table platform, and a sampling coil attached to the moving platform that provides a signal proportional to the platform velocity. An oscillator provides the driving signal for the shake table amplifier; feedback from the velocity sampling coil stabilizes the system.

The proportionality factor between the sample coil voltage and platform velocity was used to calibrate the table by three independent methods to within about $\pm$ 1.5 percent of the average value of the proportionality factor. The three methods were:

1. Honnell's method of calibration against the vertical acceleration of gravity.

2. Optical platform displacement under dynamic conditions.

3. Velocity measurements using a General Radio calibrated vibration pickup and sound-level meter.

Seismometer voltage was plotted as a function of frequency for a constant shake table velocity, typically 0.01 inch per second peak-to-peak.

During testing within the mine, the seismometer outputs were fed into an amplifier and then into a magnetic tape recorder. On playback, the recorder output was fed into 1/3-octave filters for frequency analysis. The seismometer output voltage was determined by adding the recorder level in decibels below 1 volt rms, the amplifier gain (in decibels), the playback gain, and a spectrum-level correction factor for normalizing the 1/3-octave filters to a 1-Hertz bandwidth.

The particle velocity was then determined by comparison of the calculated seismometer voltage with the response for the same frequency and a standard test velocity on the shake table. The unknown particle velocity is given by the ratio:

$$\frac{\text{Peak-to-peak particle velocity}}{\text{rms seismometer voltage}} =$$

$$\frac{\text{peak-to-peak shake table velocity}}{\text{rms seismometer voltage on shake table}}$$

The acceleration level was determined from the velocity by assuming the ground displacement to be a sinusoidal function of time. By Fourier analysis, any arbitrary waveform may be written as a sum of sines and/or cosines:

$$y = y_o \sin 2\pi f t \quad (1)$$

$$v = y_o\, 2\pi f \cos 2\pi f t \quad (2)$$

$$a = y_o\, (2\pi f)^2 \sin 2\pi f t \quad (3)$$

where:

y = ground displacement

y_o = peak amplitude

v = velocity

a = acceleration

f = frequency

Writing the corresponding time-independent terms with capital letters gives:

$$Y = Y_o \quad (4)$$

$$V = 2\pi f Y \quad (5)$$

$$A = (2\pi f)^2 \; Y = 2\pi f V \quad (6)$$

If acceleration is expressed in multiples of acceleration due to gravity (g = 978 cm/sec^2), then:

$$\text{acceleration (g's)} = \frac{A}{g} = \frac{2\pi}{978} fV$$

$$= 6.42 \times 10^{-3} fV \quad (7)$$

Background noise levels were recorded:

1. During the day when the mine was in full operation and the surface cultural noise was maximum.

2. During salt blasting operations, which occur each day at about 3:30 p. m.

3. At night, when mine operation and surface cultural noise are minimum.

Two earthquakes during testing provided a fourth background noise condition. Background noise levels during a 2-minute interval on 4 August 1966 and during blasting operations on the same day and the following day were analyzed in detail (Figures 3, 4, and 5). In reducing the remaining data, the tapes were scanned and 12 minute samples selected at random for magnitude and frequency analysis.

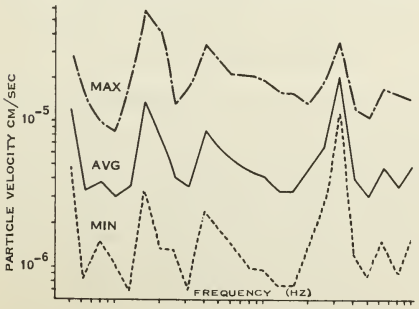


Figure 3. Background Noise (Vertical)

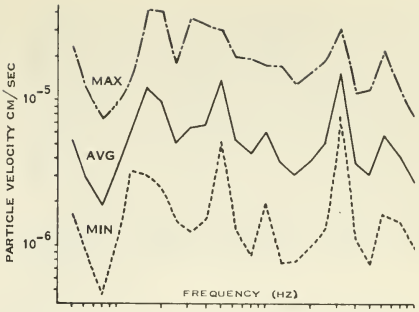


Figure 4. Background Noise (North)

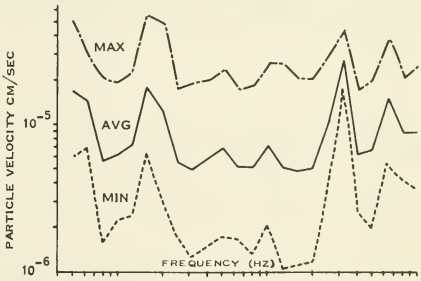


Figure 5. Background Noise (East)

The results of the analysis are presented in Table 1, which shows the highest vibration levels and the corresponding frequencies. The highest level recorded was 6.2 micro-g's at 31.5 cps. All the large-g inputs occurred at 60 cps, or very close to one of its subharmonics, and were later found to be of electrical rather than seismic origin.

FREQUENCY (HZ)	NOISE LEVEL (G'S)		
	VERTICAL AXIS	NORTH-SOUTH AXIS	EAST-WEST AXIS
1.6	1.9×10^{-7}	1.2×10^{-7}	1.4×10^{-7}
4-5	2.3×10^{-7}	4.2×10^{-7}	2.2×10^{-7}
10	4.8×10^{-7}	3.9×10^{-7}	2.7×10^{-7}
31.5	6.2×10^{-6}	3.1×10^{-6}	4.1×10^{-6}

Table 1. Detailed Analysis of 2-Minute Period

A summary of the background noise for the entire test period (4 through 8 August) is presented in Tables 2 and 3. The background noise level did not change appreciably from periods of maximum mining activity to periods of minimum activity. The average g levels measured during the 4-day period ranged from 1 to 5 micro-g's at the lower frequencies and from 10 to 30 micro-g's at the higher frequencies. The largest g level recorded, excluding mining blasts, was approximately 55 micro-g's along the vertical axis in the higher frequency region. Again, these larger magnitudes may have been of electrical rather than seismic origin. The average vibration levels (in micro-g's) along the vertical, north-south, and east-west axes over the 3-day period:

1. Vertical axis

Hertz	Micro-g's (Average)
0.5 to 5	0.37
5 to 20	4.3
20 to 200	23

AXIS	FREQUENCY BAND (HZ)	NOISE LEVEL (G'S)			
		1332 HRS TO 1938 HRS 4 AUGUST 1966	1938 HRS 4 AUGUST 1966 TO 0835 HRS 5 AUGUST 1966	2215 HRS 5 AUGUST 1966 TO 1132 HRS 6 AUGUST 1966	1135 HRS 6 AUGUST 1966 TO 2239 HRS 6 AUGUST 1966
VERTICAL	0.5–5	3.3×10^{-7} +6 DB −7 DB		4.1×10^{-7} +9 DB −6 DB	3.6×10^{-7} +11 DB −5 DB
	5–20	4.0×10^{-6} +5 1/2 DB −3 DB		4.5×10^{-6} +7 DB −3 DB	4.5×10^{-6} +6 DB −5 DB
	20–200	1.5×10^{-5} +3 DB −3 DB		2.9×10^{-5} +6 DB −1 DB	2.6×10^{-5} +5 DB −2 DB
N–S	0.5–5	2.6×10^{-7} +12 DB − 5 DB	2.6×10^{-7} +8 1/2 DB −5 DB	3.2×10^{-7} +12 DB −5 DB	2.8×10^{-7} +12 DB −6 DB
	5–20	2.6×10^{-6} +6 DB −2 DB	2.5×10^{-6} +8 DB −4 DB	2.5×10^{-6} +9 1/2 DB −4 DB	2.3×10^{-6} +11 1/2 DB
	20–200	1.2×10^{-5} +5 DB −2 DB	1.3×10^{-5} +7 DB −3 DB	2.1×10^{-5} +3 DB −2 DB	1.8×10^{-5} +4 DB −1 DB
É–W	0.5–5	3.3×10^{-7} +6 DB −8 DB	3.7×10^{-7} +7 DB −9 DB	3.7×10^{-7} +12 DB −7 DB	3.2×10^{-7} +8 1/2 DB −6 DB
	5–20	2.9×10^{-6} +7 DB −2 DB	2.9×10^{-6} +6 DB −5 DB	3.2×10^{-6} +6 1/2 DB −5 DB	2.5×10^{-6} +12 DB −5 DB
	20–200	1.1×10^{-5} +6 DB −3 DB	9.8×10^{-6} +8 DB −4 DB	1.2×10^{-5} +7 DB −4 DB	1.1×10^{-5} +7 DB −3 DB

Table 2. Analysis of Background Noise

FREQUENCY BAND (HZ)	NOISE LEVEL (G'S)		
	VERTICAL AXIS	EAST–WEST AXIS	NORTH–SOUTH AXIS
0.5–5	AVERAGE 4.3×10^{-7} +5.9 DB −2.9 DB	3.6×10^{-7} +5.8 DB −4.0 DB	3.6×10^{-7} +6.9 DB −3.9 DB
	MAXIMUM +10.6 DB	+9.6 DB	+10.3 DB
	MINIMUM −6.4 DB	−4.9 DB	−6.7 DB
5–20	AVERAGE 4.5×10^{-6} +5.0 DB −2.4 DB	2.8×10^{-6} +5 DB −2.3 DB	2.6×10^{-6} +4.4 DB −2.3 DB
	MAXIMUM +11.1 DB	+8.2 DB	+7.8 DB
	MINIMUM −2.9 DB	−3.8 DB	−3.2 DB
20–200	AVERAGE 2.9×10^{-5} +2.5 DB −1.3 DB	1.3×10^{-5} +3.3 DB −2.1 DB	2.0×10^{-5} +2.3 DB −1.2 DB
	MAXIMUM +5 DB	+4.8 DB	+4.6 DB
	MINIMUM −2 DB	−3.2 DB	−1.4 DB

Table 3. Analysis of Twelve 22-Minute Sample Periods

2. North-south axis

Hertz	Micro g's (Average)
0.5 to 5	0.28
5 to 20	2.5
20 to 200	17

3. East-west axis

Hertz	Micro g's (Average)
0.5 to 5	0.35
5 to 20	2.9
20 to 200	11.3

Hopefully, future tests will show these levels to be maintained or even substantially decreased.

The data obtained during the daily mine blasting (Table 4) show vibration levels as high as 4, 600 micro g's. This should not affect testing because the blasting time will always be known and the increased g level should not damage equipment.

DATE	FREQUENCY (HZ)	ACCELERATION (G-LEVEL)		
		VERTICAL	EAST-WEST	NORTH-SOUTH
4 AUGUST 1966	5	—	1.9×10^{-5}	—
	16	3.8×10^{-4}	5.8×10^{-5}	—
	32	2.0×10^{-3}	4.8×10^{-4}	1.5×10^{-3}
5 AUGUST 1966	1	—	6.4×10^{-6}	—
	4	2.3×10^{-5}	—	—
	16	9.6×10^{-4}	1.4×10^{-4}	—
	32	4.6×10^{-3}	9.2×10^{-4}	2.6×10^{-3}

Table 4. Vibration Levels at Test Pad Site During Salt Blasts.

Ground-Level Stability

The base rock stability (angular rotation about the north-south and east-west axes) was measured with two Talyvel electronic levels. The data indicate a total level change of approximately 1.2 arc-seconds in the east-west plane during an 84 hour period (Figure 6). The only significant disturbances during the period were two earthquakes, the first of which moved the mine floor about 0.2 arc-second about the north-south and east-west axes, and the second, which lasted about 10 minutes, moved the mine floor $\pm$ 3 arc-seconds at about 5 cps.

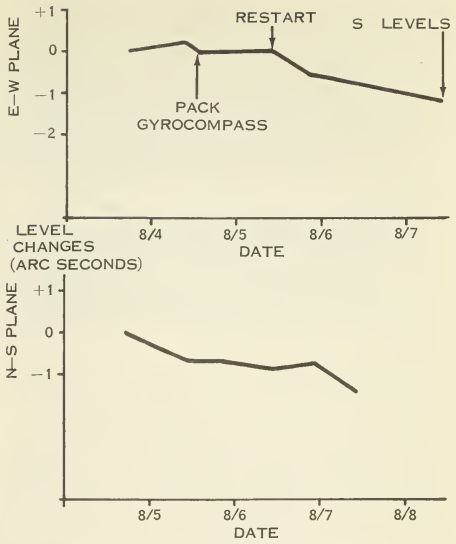


Figure 6. Summary of Talyvel Monitors

North-South Reference Line

The north-south reference line was established by an accurate single-degree-of-freedom gyro, a porro prism, an autocollimating theodolite, and an Ultradex dividing head: The gyro and porro prism were mounted in a single holding fixture, and the holding fixture was placed on the dividing head. The gyro spin axis was oriented approximately perpendicular to the plane of the porro prism and directed approximately north. The angle of deviation of the spin axis from the earth's axis in a horizontal plane was determined by taking integrated gyro torque-to-balance readings at two orientations 180 degrees apart with the input axis approximately east-west. The theodolite was used to sight in on the porro prism. The holding fixture, with porro prism and gyro, was then positioned upside down on the dividing head. The fixture was moved until the reference line between the porro prism and theodolite was again established. Gyro torque-to-balance readings were again taken, and a second angular deviation between the spin axis and earth's axis was obtained.

The average of the algebraic difference between these two angular deviations was the angular deviation between the gyro spin axis and the porro prism. With these two angles, the porro prism was moved, using the theodolite as a reference until the north-south line was perpendicular to its front surface. An optical plummet was then used to sight from the theodolite to the salt floor and a brass marker was embedded into the salt. A punch mark established the exact position on the brass marker.

This procedure was repeated approximately 34 feet from the first brass marker. The two markers, lying on the north-south line, established a semi-permanent north-south reference. This line, resighted after center punching the markers, was 4.6 arc-seconds clockwise from the desired heading. These data indicate that the desired heading was accurate to within ± 10 arc-seconds. Therefore, the reference line defined by the punch marks lies east of north (and west of south) by 4.6 ± 10 arc-seconds.

Gaseous Contaminants Concentration

A portable mine safety appliance tester was used to determine the type and concentration of gaseous contaminants in the air at the laboratory site. Hydrogen sulfide (H_2S) and carbon monoxide (CO) concentrations were well below the allowable maximum.

Corrosive Properties

To ascertain the corrosive effects of the salt mine atmosphere on metals, aluminum, stainless steel, copper, and carbon steel samples were placed in separate containers and exposed for 48 hours. The aluminum and stainless steel samples showed no visible effects. The carbon steel showed mild corrosion, while copper was heavily corroded. These conditions should not present a major problem. The air inside the laboratory will be conditioned so as to be essentially salt and moisture free.

Dust Particle Size and Count

Airborne salt and dust particles were measured and counted using a blue-grid millipore filter. A vacuum pump drew air through the filter at 1 liter per minute for 28 minutes for each of 3 samples. The average number of salt and dust particles per liter were:

Size (microns)	Particles (per liter)
0 to 10	131.9
10 to 30	50.4
30 to 50	15.8
50 to 100	3.6
100	3.0

The average number of particles per liter was 204.7.

Temperature and Relative Humidity

The dry-bulb temperature at the laboratory site is 57 degrees Fahrenheit. The relative humidity, determined from wet and dry-bulb temperature readings, is approximately 80 percent.

Safety Requirements

There are no special safety requirements except for the hard hats and safety shoes. Obviously, personnel should not wander far from the laboratory site because some panels have not been worked in 30 years and one's sense of direction is easily lost in total darkness. In the history of the mine, there has never been a cave-in.

Design and Construction of the Laboratory

The laboratory will include a clean room (100,000 particle room), storage area, and office space (Figure 7). Four pedestals extend above the floor in the clean room and one is at floor level. The three square pedestals are at working height (3 feet) for placement of dividing heads and optical equipment; the fourth pedestal, in the center, is used for mounting a porro prism (Figure 8); and on the fifth pedestal (at floor level) an air-bearing servo table is mounted (Figure 9).

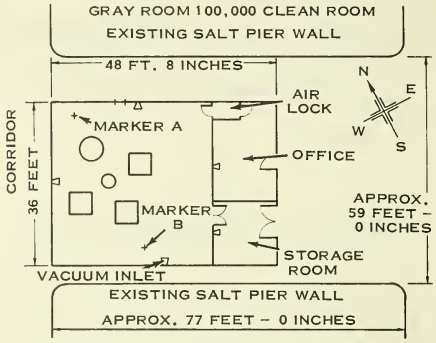


Figure 7. Laboratory Floor Plan

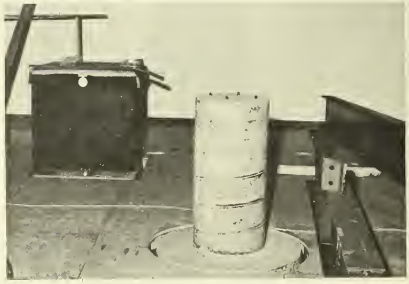


Figure 8. Pedestal

Figure 9. Servo Table Pedestal

The environmental control system consists of a continuous air-diffusing ceiling and balanced return air-wall panels. Laboratory environmental integrity, therefore, is maintained by positive air pressure in the clean room to ensure outward rather than inward air flow. The clean room enclosure wall panels consist of an exterior surface of white embossed aluminum laminated to 1/4-inch "non-com" plywood and a core of Dylite (expanded polystyrene) acrylic enameled aluminum. These materials are laminated into a modular building panel featuring lightweight, structural load-bearing capabilities and cam-locking devices. The external ceiling is made of Dylite structural panels also, but the interior surface is aluminum foil laminated to non-com plywood. The 8-foot high interior diffuser ceiling is composed of continuous supply air-diffusing panels, all suspended from the exterior roof structure. The ceiling also contains recessed rapid-start, fluorescent, troffer-type electric lighting fixtures with styrene lenses. All 4-foot panels and corner joints are air-tight, smooth, and covered with an anodized aluminum batten or cove strip (Figures 10 and 11).

Figure 10. Front View of Facility

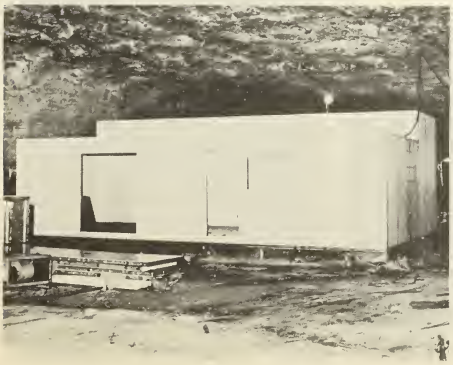


Figure 11. North View of Facility

The return air walls form an air cavity between the structural members, allowing the air to return to the recirculating fans and filtering system. All cavities have damping devices to ensure properly balanced air return from the gray room to the recirculating fans. The air-cooled, air-conditioning unit and other fan-type air-handling systems are installed on unisorb felt to minimize transmission of unwanted vibrations.

Forty-five percent efficient filters are located in the air-handling units. Eighty-five percent efficient pre-filters and 99.9 percent efficient final filters (clean room only) are located in the supply ductwork. Make-up filters are installed in the return air plenum. The air is filtered to 100,000 particles (0.5 micron and larger) per cubic foot of air. A sheet vinyl floor and central vacuum system with automatic operating controls facilitate cleaning.

A self-contained air lock and a shoe cleaner are provided at the entrance to the office area from the clean room. The air lock includes electrically interlocked doors that are automatically energized. Smocks and caps are worn. A separate air lock for material, and equipment movement will be provided, if required. Since laboratory activation, all components and systems to be tested can be handcarried. A 100-particle laminar-flow work station has been provided for inertial platform and instrument disassembly operations. This open, unencumbered work area with unrestricted freedom of movement eliminated the need for a complate class-100 clean room during initial laboratory operations.

The laboratory's primary power source for test equipment is 110/220, 60 Hertz. Line voltage regulator panels guard against electrical transients. All instrument cables from piers to consoles run through troughs under the floor so as not to interfere with operations.

Unique Construction Features

The three square test piers (Figure 12) are solid 3-foot granite cubes mounted on floor level concrete pedestals. The top surfaces of the granite blocks were lapped to a flatness of $\pm$ 0.0002 inch and leveled to within 1 arc-second of horizontal with permanent leveling jacks. All concrete pedestals were poured into holes cut approximately 2 feet into bedrock (Figures 13, 14, and 15). Before the concrete was poured, four 3/4 inch-diameter iron reinforcing rods were sunk 12 inches into holes drilled into bedrock at the bottom of each pedestal hole. The rods were set in hydraulic cement and other reinforcing material was added.

The floor of the test laboratory tests on reinforced concrete footings, 7 feet apart, poured directly on the baserock (Figures 16 and 17). A polyethylene sheet covers the floor and footings. All edges are sealed with epoxy.

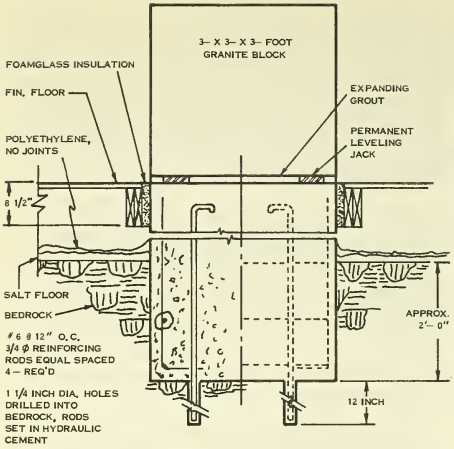


Figure 12. Typical Section - Test Piers

Figure 13. Holes in Salt and Bedrock for Pedestals

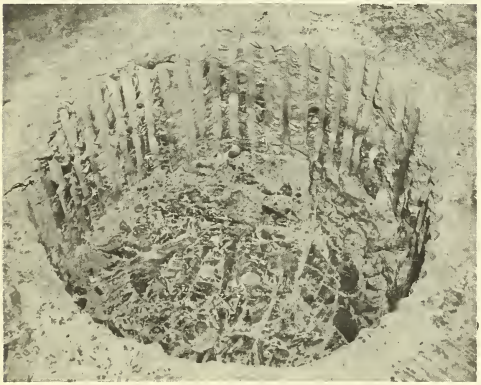


Figure 14. Detail Showing Circular Pedestal Hole

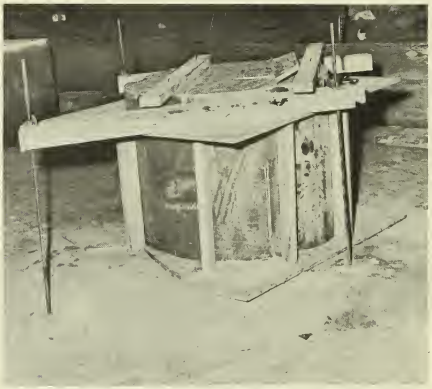


Figure 15. Circular Pedestal with Form

Figure 16. View of Footings

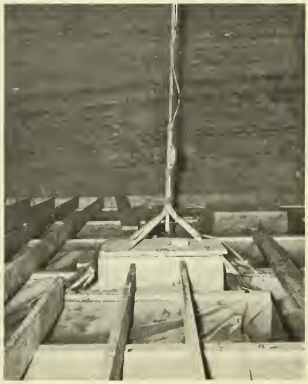


Fig. 17 View of Joists & Foundation

A 2 by 8-inch sill supports the entire prefabricated structure (Figures 18 and 19). If test equipment movement introduces vibrations to the test piers, the structure can be jacked up and insulated between the sill and the concrete footings. The test piers are isolated from the floor by foamglass insulation (Figure 9). An air space between the foamglass and the test piers minimizes transfer of energy from the floor to the piers.

Transportation of personnel and equipment, power, and all laboratory maintenance were provided by the International Salt Company. All construction labor was contracted by the International Salt Company, thus minimizing construction costs.

Laboratory construction was started in early February 1967 and completed in late May. Actual test programs were started in early June 1967. Present in-house programs will require all test stations to operate full-time through the end of the year. During testing, records of background vibration and angular tilt will be maintained to determine the effects of mine activities, local surface disturbances, earthquake disturbances, and any other cause of earth movement. If time permits, these measurements will also be recorded in other locations of the mine and at various levels.

In addition to inertial equipment testing, we expect to acquire greater knowledge of noise disturbances below the earth's surface. Our basic vibration sensing instruments will be a triad of Geo Space HS-10 seismometers arranged to provide orthogonal sensing axes: north-south, east-west, and vertical. Initially, a continuous seismological history of the laboratory site will be recorded. Subsequent vibration measurements will be monitored by threshold circuits that will automatically record any disturbances at higher amplitudes than predetermined. The three seismometers will be accurate to better than 0.1 micro-g from 1 to 100 Hertz. The angular stability of the north-south and east-west axes will be monitored by two Talyvel electronic levels accurate to less than 1 arc-second.

Figure 19. Panel at Floor Level

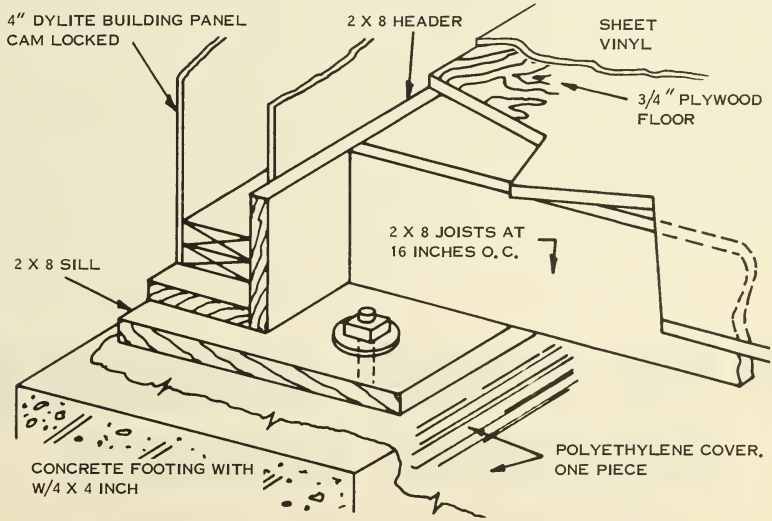


Figure 18. Laboratory Floor

67-538

A more accurate north-south line was determined before testing by gyrocompassing methods. The north-south line markers were imbedded in bedrock and aligned north-south within 5 to 10 arc-seconds. The alignment of the pedestal-mounted porro prism and the air-bearing servo table were also completed. This alignment will be updated periodically.

Inertial Instrument Test Equipment

All special and general purpose inertial test equipment presently required has been delivered. Among these are:

1. A Goerz Optical Model 577 table (with options), which is an air-bearing servo or "D" table having a ± 0.5 arc-second table axis to tilt axis orthogonality; ± 1 arc-second tilt axis readout resolution; and ± 2 arc-second tilt axis readout accuracy.

2. An Opto-metric Rotary Inclinable Leitz/Imperial optical table having a rotating table accuracy of ± 1 arc-second and tilt accuracy of ± 3 arc-seconds.

3. An Engis Model TA-52, dual axis, photoelectric universal autocollimator accurate to 0.5 arc-second.

4. An Engis Model TA-82, dual axis (coplanar), automatic-position sensing autocollimator having a direct reading to 1 arc-second.

5. Late-model power supplies, frequency and time standards, voltmeters, oscilloscopes, etc.

All test equipment was specified to general purpose requirements, but was rack mounted if possible. Gyro, accelerometer, and system test consoles were designed and fabricated by MSD-Michigan. The only special-purpose equipment required for testing was individual test panels designed to mate the specific instrument to be tested to the test console.

Conclusion

Now, when we say, "I'll see you in the salt mine," we mean it. Our way-out idea has proved not only feasible but also quite practical. We have effectively reduced such problems as:

1. Tilt inputs due to solar heating.

2. Test pad vibration due to cultural disturbances.

3. Normally high construction costs.

The facility is now in full operation. We plan to maintain permanent records of all tests conducted in the mine and to prepare a paper detailing our experiences and the knowledge gained. We shall consider any requests for admittance to the laboratory or for testing in our underground facility.

No. 67-539

STRAIN STEPS ASSOCIATED WITH EARTHQUAKES

by

C. J. WIDEMAN and M. W. MAJOR
Colorado School of Mines
Golden, Colorado

AIAA Paper
No. 67-539

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

6.04, 11.01

-- NOTES --

STRAIN STEPS ASSOCIATED WITH EARTHQUAKES*

C. J. Wideman**, and M. W. Major
Colorado School of Mines
Golden, Colorado

Abstract

The effects of earthquakes on man-made structures, such as test pads, can be of prime importance. However, it has been customary to study only the effects of oscillatory ground motions. Another type of ground motion, permanent relative-displacements, are associated with the occurrence of earthquakes. By using strain seismometers, it is possible to develop empirical relationships between earthquake magnitude, epicentral distance, and strain step amplitude. The speed of propagation of strain steps is determined to be nearly constant over a wide range of epicentral distances. The amplitude dependence upon distance is approximately $R^{-3/2}$.

1. Introduction

The effects of earthquakes on man-made structures, such as test pads, can be of prime importance. Most earthquake damage is related to oscillatory ground motion. However, there is another type of ground motion which should be recognized as being of fundamental importance, i.e., the permanent deformations resulting from strain field adjustments.

Although the seismicity of North America is generally restricted to certain zones of activity, there is no region which is truly "safe" from the effects of earthquakes. The strain field adjustments at the time of the earthquakes are not localized to the epicentral region. They extend to great distances over the surface of the earth.

Permanent adjustments in the strain field of the earth, or strain steps, following the occurrence of several large earthquakes, have been reported in the literature. However, the validity of these observations has been questioned. Press (1965, p. 2408) showed that the distant strain fields are large enough to be detected by modern instruments, and that "small changes in the length and orientation of the fault lead to order of magnitude changes in the computed strain response."

This paper will present new information on strain steps associated with earthquakes ranging in magnitude from 3.0 to 8.5. Strain steps are shown to be a real and important feature on seismic strain records. Several empirical relationships are derived which support the reality of strain steps and indicate that they are not the result of instrumental hysteresis. The amplitude of the strain steps is related to earthquake magnitude and epicentral distance. The speed of propagation for a number of the strain steps is determined.

The amplitude dependence upon distance is found to be approximately $R^{-3/2}$. The speeds of propagation and the amplitude decay with distance are used to illustrate the similarity between the mode of propagation of strain steps and surface waves.

The data used in this work have been obtained from previously published papers dealing with residual strains and relative displacements, and from the strain-meter records of the Cecil H. Green Geophysical Observatory. The observatory is near Bergen Park, Colorado, and is a part of the World Wide Network of Standard Seismographs. The station abbreviation is GOL. The geographic coordinates are 39° 42" 01' N Latitude, 105° 22" 16' W Longitude, and the elevation, h, is + 7680 ft.

Figure 1 is a map of the observatory which shows the locations and azimuths of the strain meters. The two underground strain meters were completed in February 1964, and are similar to those described by Benioff (1959). The third strain meter, which is thermally compensated, is located in the shallow trench. This instrument was completed in September 1965. The strain meters are used to measure earth tides and seismic waves.

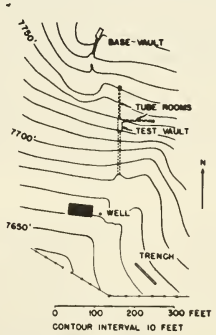


Figure 1. Map of the Green Observatory

Since the tidal strain is large compared to most seismic signals, it is not practicable to display the entire transducer output on a single record running at a gain level appropriate for seismic strains. Dual recording of the output signal is desirable and is effected by the circuit diagrammed in Figure 2.

* This work was supported by the Environmental Science Services Administration under Contract C-176-65 (G), and the Colorado School of Mines Foundation, Inc.

** Presently, Senior Engineer, Environmental Science Technology Dept., Westinghouse Electric Corp., Boulder, Colorado.

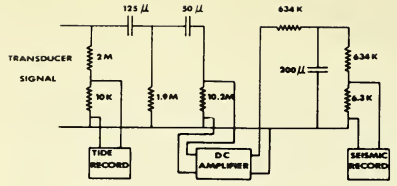


Figure 2. Strain Meter Recording System

Direct recording of the unfiltered signal at low gain on an Easterline-Angus 10-in. recorder, running at 2-in./hr., produces a record of secular strains, tides, and large earthquakes. The output of the transducer is also filtered and amplified in a method similar to that described by Major et al (1964). An Easterline-Angus 10 mv recorder writing at 15-mm/min on a paper-covered drum is used to record filtered seismic signals.

II. Strain Steps at Epicentral Distances

Permanent adjustments in the strain field of the earth are divided into two main categories for this paper. They are (1) strain steps of 10^{-8} to 10^{-10}, and (2) strain steps of 10^{-4} to 10^{-6}. The first category includes strain field adjustments which are detectable by strain meters, and the second category includes strain field adjustments which are detectable by first order triangulation.

Strain Meter Data

Figure 3 shows an example of the filtered strain records used in this paper. The earthquake which produced this record occurred on September 12, 1966, in northern California. The magnitude was about 6, and the epicentral distance was 1264 km. The step response of the filter system has been superimposed on the record. The similarity between the step response and the observed transient indicates that a strain step has occurred. A total of 22 filtered strain records with transients similar to the one shown in Figure 3 are used in this investigation.

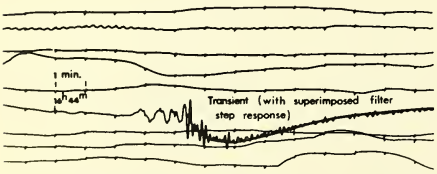


Figure 3. Record of Northern California Earthquake of September 12, 1966

Figure 4a shows three strain steps resulting from three earthquakes in the North Denver area, about 45 km from the observatory. The earthquakes range in magnitude from 4.3 to 3.5. The sizes of the transients illustrate the effect of earthquake magnitude upon the observed strain step for fixed epicentral distance. The three earthquakes occurred in a time period of one hour. The operating condition, sensitivity, and balance of the strain meter were not changed during this interval. Although the absolute value of the magnitudes of the strain steps cannot be determined, the relative sizes of the transients on the record show a positive correlation with earthquake magnitude.

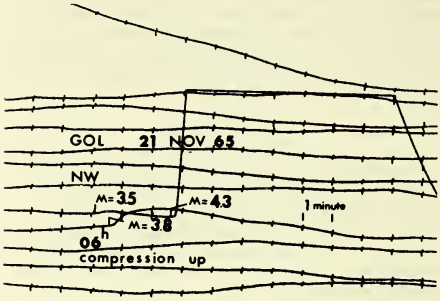


Figure 4a. Strain Meter Record of Three North Denver Earthquakes

Figure 4b is a copy of the records of the same three North Denver earthquakes obtained with Wood-Anderson seismographs. The magnitudes of these earthquakes were determined from these records. Comparison of Figures 4a and 4b shows that the transients associated with the strain steps arrive after the body phases.

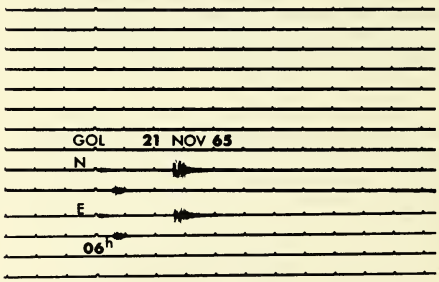


Figure 4b. Seismograph Records of Three North Denver Earthquakes

The strain step amplitudes (Figure 5) can only be estimated to an order of magnitude because precise calibrations of the strain meters were not available. The data includes strains from 10^{-8} to 10^{-10}. The empirical relationship:

$$M = 1.1 + 1.7 \log_{10} \Delta \qquad \text{eq. 1}$$

was derived from these data. Equation 1 specifies the magnitude, M, of the smallest earthquake from which strains of the order of 10^{-9} may be expected at a distance of Δ, km.

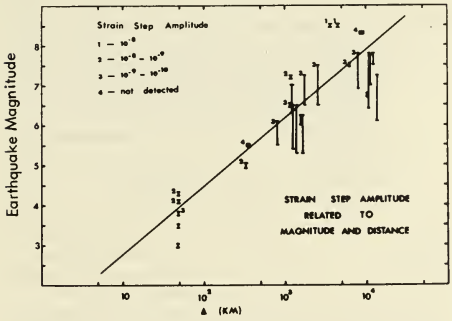


Figure 5. Strain Step Amplitude

Supplementary Data

The strain meter data reveal a dependence of strain step amplitude upon earthquake magnitude and epicentral distance. Certain other data on permanent adjustments of the Earth's strain field support this conclusion. Published reports on the relative displacements following certain earthquakes and nuclear explosions are used to estimate the strains as a function of distance within 100 km of these sources. The following information, obtained from these reports, is presented in the order of decreasing earthquake magnitudes.

The largest earthquakes studied had magnitudes near 8.5. Weertman (1965) published a report dealing with the displacements after the 1906 San Francisco earthquake. Strains of 10^{-4} and 10^{-5} were calculated from these data for epicentral distances of 5 and 30 km, respectively. The report of Press (1965) contained information dealing with the displacements following the 1964 Alaskan earthquake. Strains of 10^{-4} to 10^{-5} were calculated for epicentral distances which are similar to those obtained for the San Francisco earthquake. Strains of the order of 10^{-6} were estimated for distances of 200 to 300 km from the Alaskan earthquake.

Relative displacements following the 1927 Tango earthquake were used to infer strains near a magnitude 7.5 to 8.0 earthquake. The report of Wilson, Webb, and Hendrickson (1962) which dealt with horizontal displacements near the Gomura Fault was used to estimate strains of 10^{-4} and 10^{-5} at distances of 3 and 20 km, respectively.

The survey data of Meade (1966) following the 1966 Parkfield earthquake, magnitude about 5 3/4, have also been used to determine the strains near the epicenter. This relatively small earthquake is unusual in the fact that the fault associated with it is quite long. Strains of the order of 10^{-6} were determined at epicentral distances near 5 km.

Data which corroborate the Parkfield information is available in the report dealing with the Matsushiro earthquake zone (Kasahara and Okada, 1966). Field measurements indicated that strains of 5×10^{-5} occurred in a zone where the total seismic energy released was equivalent to a magnitude 5.8 earthquake.

Kasahara and Ikada (1966) also noted that, during the period of their observations, they felt two strong shocks. Immediate baseline resurveying showed no detectable change in baseline length. Since their instrumental limits were $\pm 2 \times 10^{-6}$, it is concluded that strains of the order of 10^{-6} do not exist at distances much greater than 1 km from earthquakes of magnitudes less than 4.3.

The Salmon seismic experiment illustrates the dimensions involved with nuclear events. The Salmon shot is equivalent to a magnitude 4.3 earthquake (Werth and Randolph, 1966). The extent of the crushed zone and plastic flow zone are of interest in the present study. Strains at a distance could not be inferred from the data because all measurements were taken within the extent of the plastic flow zone, which extended to 150 meters.

A summary of all the strain-step information, which includes both the strain meter and supplementary data, is shown in Figure 6. The data extend over 5 earthquake magnitudes, 5 orders of distance, and 6 orders of strain-step amplitude. The available data are confined to strain-step amplitudes between 10^{-4} and 10^{-10}. The largest strains, 10^{-4}, are limited by the strength of rocks, and the smallest strains, 10^{-10}, are limited by the strain-meter instrumental limits.

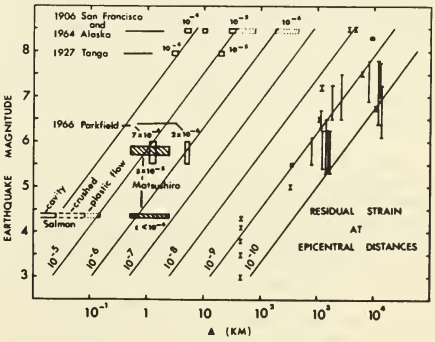


Figure 6. Residual Strain at Epicentral Distances

The information is more detailed for strain steps of 10^{-5} and 10^{-9} than for any other amplitudes. The 10^{-9} contour was obtained from the strain meter data, and is defined by equation 1. The 10^{-5} contour is obtained from the supplementary data. The slope of the 10^{-5} contour, $1.7 \log_{10}$, is the same as that for the 10^{-9} contour. The remaining contours were obtained by interpolating between 10^{-5} and 10^{-9} in a manner such that the contours were parallel and equally spaced. Although the data for strain steps of 10^{-6} to 10^{-8} are scarce, the contouring agrees well with the few observations in these ranges.

The strain field contouring of Figure 6 can be converted to rotation by assuming the measured strains to be of the shear type. The following table can be used for the conversion of strain to rotation.

Table I

Strain	Rotation in Sec of Arc
10^{-4}	1.3×10^{2}
10^{-5}	13.
10^{-6}	1.3
10^{-7}	1.3×10^{-1}
10^{-8}	1.3×10^{-2}

The rotations resulting from earthquakes are, therefore, also apparent at great distances.

III. Propagation

Two important characteristics of strain step propagation are determined for the first time. They are (1) the speed of propagation, and (2) the dependence of strain step amplitude upon distance.

Speed of Propagation

Determinations of the speeds of propagation of the strain steps have been made by using the filtered strain records. For most of the events studied, the speed is 3.0 ± 0.3 km/sec. The travel paths associated with these events are predominantly continental. The speeds determined for the remainder of the events are near 3.6 km/sec. The travel paths for these events were predominantly oceanic.

Pekeris (1955) has derived expressions for the horizontal and vertical components of displacements of a uniform elastic half-space due to the application at the surface of a point source pulse varying in time like the Heaviside unit function H(t). He shows that both of the components become "infinite" at the time of the Rayleigh wave, and that "subsequently" they revert to steady-state values. The time of arrival of the observed strain steps is in qualitative agreement with this predicted value since all the strain steps occurred after the time of arrival of the Rayleigh wave.

An unusual record was obtained from the Solomon Island Earthquake of June 15, 1966. The record shows two transients. The first transient on the record is associated with a strain step traveling at a speed of 3.6 km/sec. The travel path is predominantly oceanic so that this speed is expected. The anomalous feature on the record is the second transient. The strain step associated with this transient has a speed of 3.3 km/sec. A total of three such anomalous records were found. At the present time the arrival of two strain steps from one earthquake cannot be explained.

However, it is possible to make two conclusions about the speed of propagation of strain steps. They are:

1. For continental travel paths the speeds of propagation are 3.0 ± 0.3 km/sec. The speeds are very similar for events which originated in North Denver and for events which originated in California, Washington, and Mexico.
2. For oceanic travel paths the speeds of propagation are near 3.6 km/sec.

The velocities determined for the strain steps are very similar over a wide range of epicentral distances. They also appear to occur after the principle ground accelerations, which occur with the body phases. These two observations are considered to be additional evidence that the strain steps are real and not the result of instrumental hysteresis.

Amplitude versus Distance

Figure 7 shows the amplitudes of the strain steps as a function of distance, R. The amplitude dependence upon distance was obtained by taking a cross section of Figure 6 for a magnitude 7 earthquake. Any other earthquake magnitude could have been chosen to obtain this result. A scaling factor would be included, but the slope of the line for the observed values as plotted in Figure 7 would remain unchanged. Included in the figure are curves for amplitude dependence as functions of R^{-1}, $R^{-3/2}$ and R^{-2}. The observed amplitude dependence is seen to be very similar to that calculated for dependence like $R^{-3/2}$.

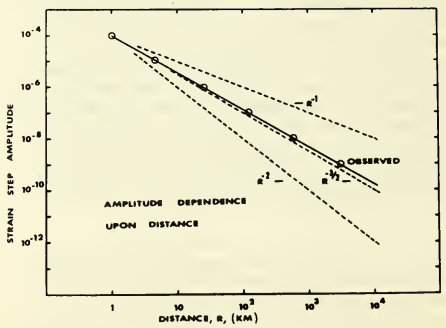


Figure 7. Amplitude Dependence upon Distance

Certain pertinent theoretical studies have been made by other investigators employing as a model a uniform elastic half-space. Dislocation theory has been used by Chinnery (1961) and Press (1965) to determine the static displacement fields caused by earthquakes. These papers show that the steady-state displacements should vary as R^{-1}. The steady-state displacements which Pekeris (1955) predicts, arriving "subsequent" to the Rayleigh wave, also vary with distance like R^{-1}. When the displacement field varies like R^{-1}, then the strain field varies like R^{-2}. This result is not in agreement with the observations shown in Figure 7.

Press (1965) accounted for certain discrepancies between observations and theoretical predictions by assuming different source mechanisms. It is well established that the source characteristics are important; however, Figure 7 shows that a large systematic error is produced whenever a model is used in which the steady-state strains vary with distance like R^{-2}. This systematic error is probably due to the mode of propagation and is not a function of the source mechanism. Therefore, no amount of adjustment of the source mechanism can correct for discrepancies found for previous models.

Surface Wave Similarities

The amplitudes of the displacements associated with surface waves characteristically vary with distance as $R^{-1/2}$. The displacements associated with strain steps should also vary with distance as $R^{-1/2}$ due to the observed strain amplitude dependence upon distance of $R^{-3/2}$. The strain steps exhibit another characteristic similar to those of surface waves: the speed of propagation. The speed is nearly constant over a wide range of epicentral distances and can be identified with continental and oceanic travel paths.

IV. Conclusions

This paper presents new information on strain steps associated with earthquakes ranging in magnitude 3.0 to 8.5. Strain steps have been shown to be a real and important feature on seismic strain records. It is concluded that:

1. The magnitude, M, of the smallest earthquake from which a strain of the order of 10^{-9} may be expected at a distance of Δ km is:

$$M = 1.1 + 1.7 \log_{10} \Delta .$$

2. The speeds of propagation of the strain steps are found to be 3.0 ± 0.3 km/sec for continental travel paths, and near 3.6 km/sec for predominantly oceanic travel paths.
3. Strain step amplitude varies with distance approximately as $R^{-3/2}$.
4. Strain steps have characteristics which are similar to those known to be associated with surface waves. Specifically they are (a) a nearly constant speed of propagation over a wide range of epicentral distances, and (b) amplitude dependence upon distance approximately as $R^{-3/2}$.

References

1. Bath, M., and H. Benioff, 1958. "The Aftershock Sequence of the Kamchatka Earthquake of November 4, 1952", Bull. Seism. Soc. Am. 48: 1-15
2. Benioff, H., 1959. "Fuzed Quartz Extensometer for Secular, Tidal, and Seismic Strains", Geol. Soc. Am. Bull. 70: 1019-1032
3. Benioff, H., 1963. "Source Wave Forms of Three Earthquakes", Bull. Seism. Soc. Am. 50: 893-904.
4. Ben-Menahem, A., and N. Toksoz, 1963. "Source-Mechanism from Spectra of Long-Period Seismic Surface Waves. 3. The Alaska Earthquake of July 10, 1958", Bull. Seism. Soc. Am. 53: 905-919
5. Chinnery, M.A., 1961. "The Deformation of the Ground Around Surface Faults", Bull. Seism. Soc. Am. 51: 355
6. Earthquake Prediction, 1965. "A Report of the Ad Hoc Panel on Earthquake Prediction", Office of Science and Technology, Washington, D. C.
7. Evison, F. F., 1963. "Earthquake and Faults", Bull. Seism. Soc. Am. 53: 873
8. Kasahara, K., and A. Okada, 1966. "Electro-Optical Measurement of Horizontal Strains Accumulating in the Swarm Earthquake Area", Bull. Earthquake Res. Inst. 45: 335-350
9. Major, M. W., G. H. Sutton, J. Oliver, and R. Metsger, 1964. "On Elastic Strain of the Earth in the Period Range 5 Seconds to 100 Hours", Bull. Seism. Soc. Am. 54: 295-346
10. Meade, B. K., 1966. Report on results of triangulation for earth movement study, vicinity of Chalome, California: Environmental Science Services Administration

11. Parkfield, 1966. "A Preliminary Report", Bull. Seism. Soc. Am. 56: 961-971

12. Press, F., 1965. "Displacements, Strains, and Tilts at Teleseismic Distances", J. Geophy Res. 70: 2395-2412

13. Press, F., and W. F. Brace, 1966. "Earthquake Prediction", Science, Vol. 152, No. 3729

14. Richter, C. F., 1958. Elementary Seismology (W. H. Freeman and Co., San Francisco)

15. Tocher, D., 1958. "Earthquake Energy and Ground Breakage", Bull. Seism. Soc. Am. 48: 147-153

16. Weertman, J., 1965. "Relationship Between Displacements on a Free Surface and the Stress on a Fault", Bull. Seism. Soc. Am. 55: 945

17. Werth, G., and P. Randolph, 1966. "The Salmon Seismic Experiment", J. Geophy. Res. 71: 3405-3413

18. Wilson, B. W., L. M. Webb, and J. A. Hendrickson, 1962. "The Nature of Tsunamis Their Generation and Dispersion in Water of Finite Depth", NESCO Technical Report, No. SN 57-2

No. 67-540

THE ALIGNMENT OF THE PRIMARY NORTH REFERENCES AT NEWARK AIR FORCE STATION

by

JOHN B. ADCOCK
Newark Air Force Station
Newark, Ohio

AIAA Paper
No. 67-540

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

-- NOTES --

THE ALIGNMENT OF THE PRIMARY NORTH REFERENCES AT NEWARK AIR FORCE STATION

John B. Adcock
Calibration and Metrology Division
Advanced Weapons Standards Branch
Newark Air Force Station
Newark, Ohio

Abstract

This paper describes the techniques which are used at NAFS to establish and maintain precisely known azimuth references in support of the present in-house inertial guidance systems and associated component repair and testing. The integral concrete piers, both internal and external to the main building, are designed to be isolated from their surrounding structure and serve as stable mountings for the necessary optical equipment. Massive structures such as the underground laboratory do not necessarily maintain the long term azimuth stability which is required. As a result the azimuth headings of the internal references are periodically verified by observations of Polaris and transfers from a 6600-foot baseline of known heading. One such internal primary north reference is located in the third level of the subterranean Calibration Laboratory and is established and maintained by means of a specially constructed slant range tunnel some 400' in length. An additional primary north reference is located in the central part of the main building. The continued verification of this azimuth accuracy presents some unique challenges and requires highly sophisticated techniques.

Introduction

The nature of the inertial guidance test and repair activity requires highly stable and precisely known internal azimuth references to support approximately one hundred test stations which must be accurately aligned with reference to the polar axis of the earth. This reference is expressed in terms of precision latitude and astronomic north.

To satisfy the requirement, it has been necessary to establish and maintain a first order capability to assure the requisite accuracy.

This paper describes the design concepts of our azimuth reference system and some of the techniques which we employ at this installation.

Site Description

Newark Air Force Station, figure 1, is located in the central section of Ohio, about two miles south of Newark in a small community known as Heath, and is approximately 30 miles due east of the large industrial complex of the state capitol, Columbus, Ohio.

The station consists of a large industrial building; an administration building; general purpose buildings; and a complete electrical sub-station. The total area consists of 38 acres and is completely fenced for security with access gates and guard posts.

The main building is a modern steel-concrete industrial structure containing 400,000 square feet of open production area which has been modified to include environmentally controlled cleanrooms which are utilized for repair, test and calibration of inertial guidance systems and components.

Since the main building was originally designed to house large capacity forging presses, many underground pits were initially constructed in the interior of the plant. The two larger pits extended sixty-five feet below ground level and have been extensively modified to serve as the home of the USAF Calibration and Metrology Laboratories.

Astronomic and Geodetic Surveys

The Coast and Geodetic Survey performed a first-order determination of the geodetic and astronomic latitude and longitude of three points at this station. A first-order astronomic azimuth determination was also made of our long baseline. This azimuth was then transferred by the C & GS to our primary reference collimators located in the main building. This transfer was performed to (a) Demonstrate that a first order azimuth could be transferred to and stored at our primary references, (b) Provide us with good initial azimuth values at our references and (c) Demonstrate short term stability of our azimuth references. The possible error assigned to each of these two indoor azimuths by the C & GS was 0.23 and 0.27 arc seconds.

When the underground laboratory was constructed provisions were made for two primary reference collimators to be mounted on shelves rigidly mounted to the exterior laboratory walls which average 6' in thickness at that depth. One reference was located on a south wall and the other on a west wall. It is noteworthy that when the collimators were aligned by C & GS, the west collimator exhibited a peak to peak azimuth change of 0.92 arc seconds while the south collimator had an oscillatory motion with a peak to peak change of 3.76 arc seconds over several days. The cause of the instability could not be determined and this reference was abandoned.

Polaris Observations

The all-hour direction method, as described in the Manual of Geodetic Astronomy, Special Publication No. 237 of the Coast and Geodetic Survey, is used to measure astronomic azimuth at this facility. This method consists of making sixteen pairs of forward and reverse theodolite angle measurements between Polaris and the distant external reference mark on each of two nights. The first order azimuth determination is based upon the mean value of at least 24 acceptable observations, not less than 12 of them on any one night, whose probable error should not exceed 0.30 arc seconds. The computation for a set of sixteen Polaris observations are lengthy and subject to computational errors, therefore, a computer program is utilized to reduce the data.

During the early attempts to make azimuth observations, one of the weakest links in the program was the real time requirement of pointing on Polaris. Since then, errors due to timekeeping instrumentation have been satisfactorily minimized by using a crystal chronometer. Figure 2 shows a block diagram of the time recording instrumentation which is presently used in the azimuth determination program.

The task of making azimuth observations from Polaris to the distant external reference mark introduces a challenge to operators who are accustomed to making horizontal angle measurements between fixed targets. Not only are they confronted with a moving target but also one that is at a considerable altitude. As a result, the choice of the method, the perfection of the instruments and the skill of the observer enter directly into the precision of the azimuth measurement.

Figure 3 summarizes the latest astronomic measurements made in determining the azimuth heading of our long baseline.

Azimuth Reference Construction

First order astronomic observations are made from the west pier system as shown in figure 4 for determining the azimuth angle of the distant reference mark. The reference mark, figure 5, is located 6600 feet north of the installation and consists of a standard brass disc set in a square concrete monument that is flush with the ground. A portable target light assembly is used to illuminate the reference mark during the night observation program. There are two methods by which the target light can be aligned with the reference mark. By the first method, the target light is extended above the reference mark by means of a tripod. The alignment process is completed by optical means thereby assuring the accurate placement of the target light with respect to the reference mark, however, it requires considerable setup time. By the second method the target light is placed directly on the reference mark and aligned visually. This can be repeatedly done without introducing any large errors in the observation program for at the distance of 6600 feet an angle of 1 arc second represents a displacement of approximately 0.38 inches.

The west pier, figure 6 and 7, is a steel reinforced concrete structure which extends five feet below the existing floor to a footer resting on undisturbed earth. The pier is isolated from the surrounding structure by means of a soft mastic joint. A plate nine inches in diameter is affixed to the top surface of the pier for repeatable positioning of the theodolite. The west pier and associated equipment is protected from the weather and sun effects by means of a wooden structure, one side of which opens to permit observations.

The north pier, figure 8, is located within the main building at a distance of 1025 feet from the west pier. The construction of the north pier is identical to that of the west pier except for its height above the main floor. The pier provides a stable mount for the target collimator which serves as the internal azimuth reference for the specialized repair and test areas at main floor level. The collimator is simply a telescope focused carefully for parallel light which has a suitable illuminated reticle. Both pier and reference are protected against possible personnel disturbances by means of a steel and plexiglass housing attached to the main floor.

In order to verify the azimuth heading of the north pier reference, a traverse is required at the east pier within the main building. The east pier, figure 9, is constructed similarly to the previously described west pier. A theodolite mounting plate is affixed to the top surface of the pier for repeatable positioning of the instrument. A 1/10 wave reference mirror assembly, rigidly mounted to the pier, is utilized in the azimuth measurement program.

An additional internal azimuth reference is located in the third level of the Calibration and Metrology Division subterranean laboratory approximately 35 feet below the main floor level. The azimuth reference, figure 10, is mounted on a shelf which is rigidly attached to the external laboratory walls. A traverse is required from the lab pier in order to maintain the azimuth reference. The lab pier is attached to the laboratory floor which is a solid mass of steel-reinforced concrete some 24 feet in thickness. Line-of-sight between the external and internal reference is obtained by using a special slant range tunnel which will be discussed in the following paragraphs.

Slant Range Tunnel

The problem of establishing and maintaining a precise azimuth reference in the subterranean laboratory was recognized during the early stages of the planning program. Since the laboratory was to be used

to evaluate inertial sensors under sophisticated environmental conditions, the maximum tolerable error of ± 2 arc seconds was established for the azimuth reference.

Several approaches were thoroughly investigated by government agencies, as well as professional consultants, to support the azimuth requirements of the laboratory. One approach required an access through the building structure by means of a vertical tube extending from the laboratory working level through the intervening floors and building roof. Observations would be made from a roof mounted movable turret and in turn transferred to the laboratory level by optical means. This method was determined to be unsatisfactory for the following reasons:

(1) The problems associated with satisfactory insulation of such a tube, at least 120 feet long. The effects of temperature gradients along such a length would negate the required precision of the associated optical system.

(2) The dynamic stability requirements for the roof mounted turret; in a structure such as was planned, and of the height involved, the system would be subjected to the constant expansion and contraction characteristics of the main building, therefore precluding accurate positioning within permissible tolerances.

(3) Refraction effects associated with observations above the roof: the vast area of the roof, a total of approximately 9 acres, and the refractions associated with it would introduce intolerable error sources into the measured values.

(4) Optical transfers with the system: the number and type of transfers through such a system involving critical alignment of the optical elements would compromise the accuracy beyond that which would be tolerable in addition to the problem of initial position verification, which must, of necessity be accomplished.

(5) Cumulative adverse effects: the errors in such a system, along with the associated variables, would incorporate indefinable uncertainties.

Other approaches were reviewed and rejected - among them were:

(1) Direct observation of Polaris from the laboratory level: this approach was deemed to be impractical due to the laboratory wall thickness, the underground location and the intervening structures within the main building.

(2) The use of two tubes to transfer an external reference over the top of the laboratory and thence vertically downward to the required level: this method was considered very promising until Coast and Geodetic Survey analyzed the situation and advised that an accuracy no better than ± 5 arc seconds could be achieved by this method.

(3) The use of an electro-optical system consisting of a beam of polarized light and a quartz crystal analyzer to transfer a reference vertically to the laboratory level: the system could only be proven under simulated conditions otherwise the system would be vulnerable to undetectable error sources during and after installation.

As a result of the investigation of all approaches, the slant range tunnel was the only method which could be considered as practical. The technique was reviewed by numerous professional personnel who agreed that other approaches would generate more serious problems. By using this particular method the cumulative errors introduced by additional transfers from the external baseline to the laboratory azimuth reference would be minimized.

The slant range tunnel extends 400 feet from the external west pier system to the laboratory interior as shown in figure 11. The tunnel consists of welded sections of 12-3/4 inch O.D. aluminum pipe that is supported by roller assemblies. Above ground level, the tunnel is wrapped with insulation; below ground level the tunnel is enclosed in a 30 inch casing for access to the lower roller assemblies should realignment of the tunnel be required in future years.

Holding fixtures are provided at the ends of the tunnel for mounting the special 10 inch optical flats. An adjustable turret assembly at the tunnel entrance provides a means for adjusting the upper flat in elevation and azimuth. The tunnel is airtight and is continually evacuated to a pressure of about 1 millimeter of mercury to assure that thermal gradients within the system will not deviate the beam of light systematically during the observation program. Whenever it is required to purge the tunnel, a supply of nitrogen is used to eliminate the possibility of foreign particles being injected into the system.

For high order accuracy the faces of the individual flats should be parallel to within 0.1 arc second and flat to 1/10th wave. An alternative to this approach, using flats of lesser quality, would be to have the capability of rotating each flat 180° about the tunnel axis and using only the center part of the lens. By taking half the readings through the tunnel with the flats in one position and the other half with the flats rotated 180°, the effect of the residual non-parallelism can be balanced out in the observing program. This approach, however, compounds the evacuation process.

Azimuth Reference Measurement

When the azimuth heading of the distant external mark has been established, the next

task is to tie this line to the internal reference collimators with minimal accuracy deterioration.

To verify the azimuth heading of the reference collimator located in the Calibration Laboratory, it is necessary that the two optical flats on the tunnel ends be reasonably parallel to each other since an angle of 1 degree between the flats would cause a deflection in the beam passing through the tunnel of about 1 arc second.(3) The flats can be readily aligned to within 10 arc seconds by the following methods.

(1) By autocollimation the lower theodolite is aligned normal to the lower flat.

(2) The upper theodolite at the west pier is then aligned with the lower theodolite so that their optical axes are parallel.

(3) With the position of the upper theodolite fixed, the upper flat is tilted about in azimuth and elevation until autocollimation with the upper theodolite is achieved.

Electronic consoles near the line-of-sight in the Calibration Laboratory are turned off during the observation program to minimize the effects of image deterioration caused by temperature gradients and resulting turbulence. One week of favorable weather is normally required to verify the azimuth heading of the laboratory reference collimator.

The laboratory azimuth reference was recently verified against our long baseline. The 1 sigma RMS value of $\pm$ 1.23 arc seconds between the baseline of known heading and the slant range tunnel was masked by atmospheric conditions which existed during the series of observations. The limited time requirement for a precise laboratory azimuth did not permit additional observations to further improve the accuracy of the reference. Measurements from the slant range tunnel to the azimuth reference collimator resulted in a 1 sigma RMS value of $\pm$ 0.66 arc seconds. By taking the square root of the sum of the squares the 1 sigma RMS value for the laboratory reference collimator is reduced to $\pm$ 1.6 arc seconds.(2)

The west pier is also used to verify the azimuth heading of the reference collimator located at the north pier within the main building. A precision extension plate, figure 12, designed and furnished by the Coast and Geodetic Survey, is mounted to the existing theodolite mounting plate. With the theodolite height extended, the line-of-sight to the internal east pier is obtained. To minimize the possible effects of image deterioration due to turbulence at the building entrance point, a special grid is used in place of an optical flat. The grid, figure 13, is coated with a very thin film consisting of 20 parts of clear lacquer and 1 part of tricresyl phosphate.

The line-of-sight within the main building, leading to the east pier, passes through a series of large open windows at the upper level of the Calibration Laboratory. Several cooling carts and electronic consoles are adjacent to the line-of-sight that extends to the reference collimator located at the north pier. The effects of the heat generating equipment have made it necessary to increase the number of observations in order to verify the azimuth heading of the reference collimator to the required accuracy.

Prior to the observation program, the precision theodolites at the west and east piers are carefully aligned until their axes are parallel and nearly coincident. With this line established, the reference mirror at the east pier is aligned to the internal theodolite by autocollimation. Thus, the internal theodolite provides a fixed target for the external instrument which is referenced to the initial alignment.

The reference mirror also serves as a means for determining the azimuth stability of the reference collimator during periods of inclement weather. During a recent series of measurements, consisting of 16 sets of data, a dispersion of $\pm$ 1.8 arc seconds was obtained. The measurements included the effects of temperature gradients and resulting turbulences which are present along the line-of-sight. As a result of the data obtained, the 1 sigma RMS value for the measurement is reduced to $\pm$ 0.9 arc seconds which satisfies the present azimuth stability requirements for the reference collimator.

In April 1967 a local earthquake generated additional measurements to determine if a change in the azimuth references had occurred. The internal references were verified and no measurable change in the azimuth headings of the references could be detected.

Conclusion

In order to satisfy the requirements of future inertial sensors, it has been necessary that our azimuth reference program be upgraded. To meet the new requirements a program using improved equipment and new techniques is being developed to provide azimuth references which are easily obtainable and extremely stable over long periods of time.

References

(1) Manual of Geodetic Astronomy, Special Publication No. 237, from the United States Coast and Geodetic Survey.

(2) McNish, A. G. and Cameron, J. M., Propogation of Errors in a Chain of Standards, presented at the Conference on Standards and Electronic Measurements and printed from IRE Transactions on Instrumentation, Vol. I-9, Number 2, Sept. 1960.

References (Cont'd)

(3) Jenkins, F. A. and White, H. E., Fundamentals of Optics, Mc-Graw Hill Book Company, Inc., 1957.

Figure 1 Aerial View of Facility

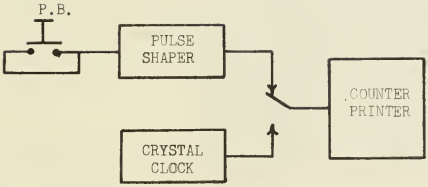


Figure 2 Diagram of Time Recording Instrumentation

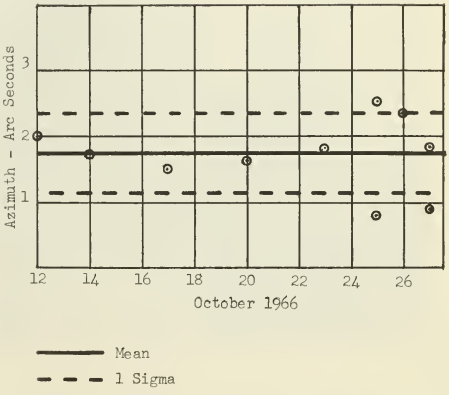


Figure 3 Baseline Azimuth Determinations

Figure 4 West Pier System

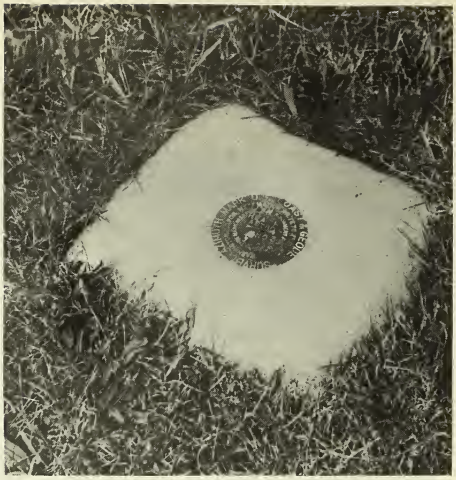


Figure 5 Distant Reference Mark

Figure 6 West Pier - Theodolite Mounted

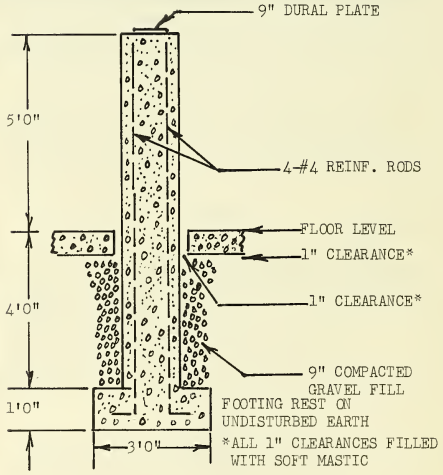


Figure 7 - West Pier Construction

Figure 8 North Pier - Collimator Mounted

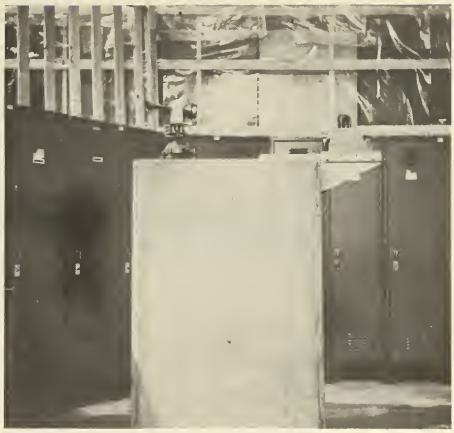


Figure 9 East Pier and Mirror Assembly

Figure 10 Laboratory Azimuth Reference Collimator

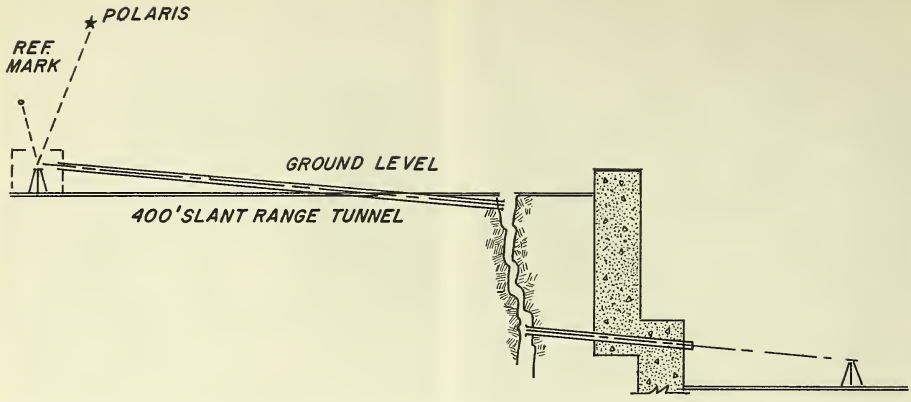


Figure 11 Drawing of Slant Range Tunnel

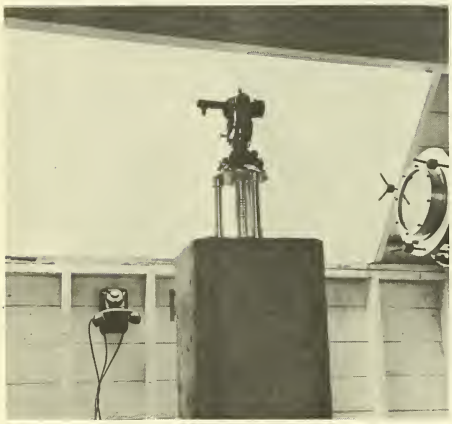


Figure 12 Theodolite Height Extension Plate

Figure 13 Optical Membrane

-- NOTES --

-- NOTES --

No. 67-548

PERFORMANCE CHARACTERISTICS OF AN AUTOMATIC PLATFORM TILT STABILIZATION AND VIBRATION ISOLATION SYSTEM

by

JEROME S. PEPI and RICHARD D. CAVANAUGH
Barry Wright Corporation
Watertown, Massachusetts

AIAA Paper
No. 67-548

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

10.05, 10.10

-- NOTES --

PERFORMANCE CHARACTERISTICS OF AN AUTOMATIC PLATFORM TILT STABILIZATION AND VIBRATION ISOLATION SYSTEM*

Jerome S. Pepi, Project Engineer
and
Richard D. Cavanaugh, Senior Staff Engineer
Barry Controls
Division of Barry Wright Corporation
Watertown, Massachusetts

Abstract

Barry Controls has recently provided the Advanced Weapons Standards Branch of the Air Force (HCLW) with a platform that is automatically tilt stabilized to within ± 0.1 seconds of arc over a 72-hour period, relative to the local gravity vector while passive azimuth control of the platform relative to the local foundation is maintained to within ± 0.25 seconds of arc. A feedback control system utilizing entirely pneumatic computing components in conjunction with pneumatic vibration isolators was employed to provide the tilt stability. This paper represents a review of the principles of operation of the system, basic design calculations and their effect on component selection, a review of test data, and a projection as to what further improvements in the state-of-the-art can be expected when utilizing these techniques.

I. Introduction

The tilt of ordinary laboratory foundations can be affected by such conditions as earthquake waves, Earth tides, thermal distortion of buildings and surrounding ground, instability of construction materials, and precession of the polar axis. Daily variations in ground tilt normally vary from ± 3 seconds of arc to ± 10 seconds of arc depending upon the geophysical location and soil conditions. Yearly variations of as much as 90 seconds of arc have been recorded.[1] The condition having the most pronounced cyclic effect on the tilt of conventional concrete foundations is the temperature gradient imposed on the surrounding building and soil while permanent tilts are primarily due to earthquake waves. Also, in addition to these low frequency factors, there are additional man-made disturbances due to traffic and industry in the frequency range above 2 Hz.

However, since precision gyroscopes and accelerometers are the basic motion sensing devices in inertial navigation systems, the instruments used in their calibration, in addition to other level sensing control systems such as those used in orbiting geographical observatories, require a stable foundation for extended periods of time. The approach employed in providing the Advanced Weapons Standards Branch of the Air Force (HCLW) with a tilt stabilized test pad is based on a unique method of pneumatic feedback control in conjunction with pneumatic vibration isolators. This application provides a test pad stability comparable in performance to any pier known to the authors (i.e., ± 0.1 seconds of arc per 72 hour period) and the techniques employed will allow further improvements in the state of the art.

The pier as supplied to HCLW has two main functions. Firstly, it provides a pier that is automatically tilt stabilized relative to the local gravity vector for test loads up to 10,000 pounds, while passive azimuth control is maintained relative to the local foundation. Since the computed daily variation in the direction of the local gravity vector is approximately ± 0.02 seconds of arc due to the effects of the moon, this technique provides on a daily basis a reference plane relative to the Earth's inertial coordinates within the present state-of-the-art in gyro drift testing. Secondly, it provides vibration isolation of the high frequency translational excitations of the foundation. The former being necessary for gyro testing and the latter being required for accelerometer testing.

A 0.5 Hz translational resonant frequency isolation system was desired in order that the vibratory accelerations of the pier would be maintained below 1×10^{-6} g. However, it also had to be relatively unresponsive to active test loads such as air bearing turntables with horizontal imbalanced moments rotating at speeds of up to 1000 earth rate. Thus, since the stiffness of a cylindrical pneumatic isolator is inversely proportional to its column height, the lower the resonant frequency of the isolation system, the slower will be the control response to the tilt excitations of the pier due to the necessity for controlling the pressure in an increased volume. The final design is, therefore, a compromise between the desired vibration isolation and the speed of response of the control system to the low frequency and transient rotational excitations. The following represents a technical review of the system and a discussion of the engineering compromises involved.

* Sponsored by the Advanced Weapons Standards Branch of the Air Force
Newark AF Station, Ohio

II. Initial Considerations

Pneumatic components were selected for both the pier position control and vibration isolation systems since they offer the following advantages:

(a) Air is a convenient industrial medium;

(b) Inherent reliability is associated with mechanical components;

(c) Pneumatic isolation systems employing relaxation damping concepts provide increased isolation efficiency over that for viscous-and Coulomb-damped isolation systems with equal resonant frequency and resonant transmissibility;

(d) The position of the pier can be controlled by varying the pressure in the isolators, thus allowing continuous vibration isolation while compensating for tilt excitations;

(e) Pneumatic isolators utilizing rolling sleeve diaphragms for the flexible element are capable of transmitting the forces necessary to support the high static loads of the pier while providing essentially infinite resolution.

Their limitations are a function of:

(a) The compressibility of air and its effect on the system response time to tilt excitations;

(b) The hysteresis in state-of-the-art pneumatic computing components is 1/8 inch of water gage.

The dynamic characteristics of a pneumatic isolation system utilizing relaxation damping principles has been extensively described in the technical literature (References 2, 3, and 4.) However, since it provides one of the primary functions of the pier, a qualitative review is appropriate.

Figure 1 illustrates the application of surge tank damping to a pneumatic isolator while Figure 2 illustrates the mathematical model of its elements. When the damping constant "c" is zero, then the resonant frequency of the isolation system is determined by the springs of rates K_1 and K_2 in series with the mass. When the damping constant is infinite, then the resonant frequency is determined by the spring of rate K_1 and the mass. In each case the high frequency attenuation rate is 12 db/octave corresponding to that for an undamped system, as illustrated in Figure 3. By varying the value of damping between the two extremes, according to definite formulas, it is possible to obtain an isolation system with an optimum (minimax) value of transmissibility at resonance which retains the high frequency attenuation characteristics associated with an infinite value of "c". The advantage of this device can be inferred from Figure 3. For a 0.5 Hz resonant frequency, greater than five times the isolation is provided over that for a viscous-damped system with equal resonant frequency and resonant transmissibility in the frequency range from 10 to 30 Hz where the predominant source of ground vibratory accelerations occur. (1)

Horizontally, the isolator stiffness is obtained from the air surrounding the elastomeric diaphragms and the shear strains induced in them due to horizontal deflections. Damping is due to the hysteresis within the diaphragms. Thus, horizontally the characteristics of a conventional spring mass system are obtained. A resonant frequency of 1.8 Hz and an equivalent viscous damping ratio was provided for this application. An added feature of this configuration is its self-centering characteristics. This principle provides excellent azimuth stability of the pier relative to the local foundation when three or more isolators are utilized.

The significant features of the pier provided to HCLW are illustrated in Figure 4. This configuration was selected since an 8-foot by 8-foot by 10-foot deep well was available for the installation. A 90 in. sq. by 1-foot thick granite plate comprises the working surface which is in turn supported by four pneumatic isolators at the corners via an intermediate substructure. The flexible elements are at the top of the isolators. Mounted to the bottom surface of the granite plate are a height sensor and an instrumentation platform which supports the level sensors and pneumatic computing components. Encompassing the instrumentation platform is a personnel access platform which is attached to the foundation.

Counterbalancing weights are supported by the isolated substructure. Also, stiffeners have been added between the isolators in order to minimize the error in azimuth position between the diaphragms and the base of the isolators. An azimuth stability of better than $\pm$ 0.25 seconds of arc relative to the local foundation was obtained.

III. Tilt Stabilization System

Figure 5 schematically represents the tilt control system of the pier. Although four isolators are used to support the platform, the system acts as if it were a three point suspension. This was accomplished by interconnecting two of the isolators on one side of the platform so that the vertical position of their flexible elements are governed by the same height sensor which is located midway between them. Separate platform level sensors mounted on the instrumentation platform, with an included angle of 90° between them, govern the air pressure in the other two isolators. This configuration effectively controls the pitch and roll attitude of the pier working surface.

The height sensor is located on the north-south centerline at the front edge of the pier beneath the working surface support. It consists of a servovalve which admits or exhausts air from the height control isolators until a preadjusted height is satisfied within $\pm$ 5 x 10^{-5} inches. When the platform is depressed below its mean position by the addition of a load, the actuating lever exerts an upward force against the sensor spool causing the supply air to be transmitted to the isolators and vice versa.

The level sensors which are located along the platform diagonals transmit pneumatic signals which are proportional to their angular deviation from the local gravity vector. This signal consists of a variable back pressure which is developed by the pendulum moving closer to or further from two nozzles (as induced by platform tilt) which are symmetrically located either side of it. Two

nozzles are used in order to balance the forces exerted by the impinging air, to minimize the effects of temperature and barometer pressure variations, and to increase gain. The nozzle flow controls are set such that when the pendulum is exactly centered between the nozzles, the nozzle back pressures are equal. The signal which is utilized by the pneumatic computing circuit to be discussed later is the difference between the nozzle back pressures.

The pendulum consists of a precisely machined bar of steel to insure accurate location of the CG along its vertical centerline. Also, it was made purposely heavy to minimize the effects of the air jets impinging upon it. The pivot point consists of a ground stainless steel knife edge resting in a V-groove. These parts were hardened to 55 on the Rockwell "C" scale in order to minimize plastic deformation at the points of contact. Although this method of pivoting the pendulum has provided an adequately friction-free suspension for this application, if further improvements in tilt stability are to be obtained, an air bearing or similar device will have to be employed as the pivot. The bottom of the pendulum rests in a high viscosity silicone oil which is used to critically damp the oscillations of the pendulum.

Temporarily neglecting the effects of the air impinging upon the pendulum, the linearized open loop pressure versus position relationship in LaPlace notation for the flapper-nozzle amplifiers illustrated in Figure 6 is given by

$$\Delta P_m(s) = \frac{2P_o \dfrac{X_n(s)}{X_o}}{1+\tau s} \qquad (1)$$

where:

$$\Delta P_m(s) = P_{m1}(s) - P_{m2}(s) \qquad (2)$$

$$P_o = \frac{A_i P_i}{\pi d_n X_o} \qquad (3)$$

$$\tau = \frac{1.3 \times 10^{-4} V_m}{\pi d_n X_o} \qquad (4)$$

for air under standard conditions. Also, the flow through the orifice and nozzle is considered to be choked. The corresponding gain can now be written as

$$\frac{\Delta P_m(s)}{X_n(s)} = \frac{2\dfrac{P_o}{X_o}}{1+\tau s} \qquad (5)$$

From equation (4) it can be seen that in order to obtain a low time constant for a fast system response, V_m should be small and $X_o d_n$ should be large. On the other hand, in order to obtain a high static gain for measurement accuracy, a low value of $X_o d_n$ is required, as can be seen from combining Equations (3) and (5). Thus, an engineering compromise must be made. For the design under consideration, the value of τ is 3 seconds. Also, P_o is 26.2 psia and X_o is 0.002 inches for a predicted static gain of 26,200 psid/inch. This agrees very closely with the experimental value of 25,000 psid/inch in the X_n range of less than $\pm$ 0.001 inches as illustrated in Figure 7. Based on a spacing of 7 inches between the nozzles and the pivot point of the pendulum, the above gain corresponds to 0.9 psid/second of arc.

When the pendulum is deflected from its mean position due to a tilt excitation of the pier, however, the differential forces which act on it due to the impinging air tends to return it to its initial position thus reducing the effective gain. Taking these effects into account, the effective differential gain of the level sensor in LaPlace notation is given by

$$G_p(s) = \frac{4.85\times10^{-6} L_n \left[\dfrac{\Delta P_m(s)}{X_n(s)}\right]}{1+\dfrac{\pi d_n^2 L_n^2}{4 W_p L_c}\left[\dfrac{\Delta P_m(s)}{X_n(s)}\right]} \qquad (6)$$

where G_p (s) is expressed in terms of psia/second of arc deflection of the pier relative to the local gravity vector. The static value of G_p (s) for the HCLW system is 0.5 psid/second of arc.

The total hysteresis of the pneumatic computing components in each of the level control circuits, which will be discussed below, is $\pm$ 1/2 inch of water gage or $\pm$ 0.018 psig. However, the computing circuit also increases the level sensor gain by a factor of 3. Thus, neglecting the hysteresis in the pendulum pivots, the maximum expected drift of the pier in either axis was $\pm$ (0.018/1.5) or $\pm$ 0.012 seconds of arc which was significantly less than the design goal of $\pm$ 0.5 seconds of arc.

The flow of air in the flapper-nozzle circuits is, however, extremely low and if they were used to directly control the flow of air into the isolators, there would be a considerable delay between pier load changes or foundation tilt and the restoration of the pier to level. In order to increase the rate of signal pressure change and the magnitude of flow due to angular displacements of the pier, a pneumatic computing circuit has been employed as illustrated in Figures 5 and 8.

The function of the computing relay is to provide the summation

$$P_c = P_f + \Delta P_m \qquad (7)$$

The weight rate of flow in the feedback circuit $\dot{W}_4$ is given by

$$\dot{W}_4 = \frac{g V_f}{kTR}\frac{dP_f}{dt} \qquad (8)$$

where the constants are defined in Figure 6. In

LaPlace notation, Equation (8) becomes

$$P_f(s) = \frac{1}{s}\left[\frac{kTR}{gV_f}\mathcal{L}\left\{\dot{W}_4\right\} + P_f(0^+)\right] \quad (9)$$

Taking the LaPlace transform of Equation (6), substituting Equation (9) into the resultant expression and then taking the inverse transform

$$P_c = P_f(0^+) + \Delta P_m + \frac{kTR}{gV_f}\int \dot{W}_4 dt \quad (10)$$

When the control system is originally turned on, $P_f(0^+)$ is zero. Referring again to Figure 8, however, a pressure booster is used to multiply the output of the computing relay by three and to increase the flow to the level control isolator. Thus, when the level sensor is nulled and awaiting rotational excitations, P_c and $P_f(0^+)$ are equal to one-third the isolator gage pressure.

The expression for $\dot{W}_4$ is the standard nonlinear orifice flow equation which is a function of P_c and the ratio P_f/P_c. Since $\dot{W}_4$ goes as $(P_c - P_f)$ which is equal to ΔP_m from Equation (6), the last term on the right side of Equation (10) is termed "integral control" while ΔP_m alone represents proportional control. The integral control provides for a steadily increasing pressure to the isolators the longer the platform is deflected, up to the limit of the supply pressure.

In order to understand the importance of integral control, consider the relation

$$C_1(\phi - \theta) + C_2\phi + C_3\int \phi dt = M \quad (11)$$

where θ and φ are the angles that the floor and pier make with the gravity vector respectively, M is the moment applied to the pier, C_1 is the rotational spring stiffness of the isolator diaphragms and C_2 and C_3 represent the proportional and integral control gains respectively. Equation (11) describes the response of the pier to tilt excitations of the foundation for frequencies less than 10^{-2} Hz and the mean position of the pier when responding to induced moments. The complete expression would include the inertia of the pier, isolator damping, and the time lag effects in the tilt control system.

Considering a step moment of M_o and a step tilt of the foundation of magnitude θ_o, Equation (11) in LaPlace notation becomes

$$\phi(s) = \frac{C_1\theta_o + M_o}{s(C_1 + C_2) + C_3} \quad (12)$$

The inverse transform of Equation (12) is

$$\phi(t) = \frac{(C_1\phi_o + M_o)}{C_1 + C_2}\exp\left[-\left(\frac{C_3}{C_1 + C_2}\right)t\right] \quad (13)$$

Thus, if proportional control only was employed (i.e., $C_3 = 0$), there would be a steady state error in the angular position of the pier depending upon the relative magnitudes of C_1 and C_2. Experiments showed that with C_3 equal to zero, the pier would sustain up to a 1 arc second shift due to the addition of a horizontal 5-foot pound moment. However, with a relatively high ratio of $[C_3/(C_1 + C_2)]$ as is employed, theoretically there will be no steady state error in the angular position of pier due to a tilt of the foundation or an applied moment. Again experimentally, it was determined that with integral control, the pier has a negligible sensitivity to a 5-foot pound moment while a 3000-foot pound moment will only cause a sustained shift of 0.5 seconds of arc.

Including the inertia effects of the pier and isolator damping, the solution of $\varphi(t)$ to step excitations of φ_o and M_o is of the form

$$\phi(t) = \theta_o\left[Ae^{-p\omega_p t}\sin\left(\omega_p\sqrt{1-p^2}\,t + \psi\right) + Be^{-q\omega_p t}\right] + \dot{M}_o\left[Ce^{-p\omega_p t}\sin\left(\omega_p\sqrt{1-p^2}\,t + \alpha\right) + De^{-q\omega_p t}\right] \quad (14)$$

A typical rotational time history due to a moment excitation is given in Figure 9 to illustrate the physical significance of the parameters. The terms containing B and D are the average values about which the terms containing A and C oscillate. The value of $q\omega_p$ determines how rapidly the average value of the rotation decreases to zero and the value of $p\omega_p$ determines how rapidly the oscillatory terms die out. It would seem then, that making both p and q as large as possible would give the best transient response.

As with most engineering problems, however, the performance of this design is a compromise. Vibration isolation requirements and the stability of the control system as well as the flow capacity of the components puts limits on both p and q. The parameters A and C are the inverse functions of $(1-p^2)$. Thus, increasing p by means of the effects of integral feedback control causes an increase in the amplitude of the oscillations although they die out faster. Increasing the value of p also degrades vibration isolation since it requires that damping be increased in order to maintain a stable control system. Also, increasing p and q requires an increase in the size of the flow control components thus in general reducing their accuracy. The final design is therefore a balance between integral control and isolator damping to give the overall results of good transient response to tilt excitations and translational vibration isolation.

Figure 10 illustrates a time history of the tilt response of the HCLW pier to a step moment of 80-foot pounds. The step moment was due to laterally shifting a 40-pound load a distance of 2 feet from the center of the pier. Figure 11 is a plot of the tilt stabilization times for various magnitudes of step moments. The constant slope for moment inputs of greater than 20-foot pounds is due to the integral feedback control which provides for a steadily increasing signal to the isolators the longer the pier is deflected, as mentioned earlier.

Thus, although one minute is required for tilt stabilization following a 30-foot pound moment, only ten minutes are required for a 1500-foot pound moment. If an operator walks a complete circuit on the periphery of the platform and then steps off, two to three minutes are required for tilt stabilization. The time increases to four minutes when three operators walk on and about the platform and then step off simultaneously.

Figures 12 and 13 illustrate time histories of the tilt position of the pier along one of its diagonal axes as compared to the tilt of the foundation. The tests were conducted at HCLW. Readout sensitivity of the tilt monitoring system which employed Ideal Aerosmith mercury pool-type manometer transducers was 0.025 seconds of arc. With this sensitivity, no transient response of the pier to the step tilts of the foundation between the elapsed time of 45 and 50 hours could be detected for the Figure 12 test.

Figure 14 is a graph of rotational transmissibility for the HCLW pier as a function of frequency which was computed assuming no hysteresis in the tilt control components. Essentially all ground rotations are isolated for frequencies less than 5×10^{-3} Hz. Above this frequency, the frequency response capabilities of the tilt control system begin to fall off until the isolation is entirely passive above 10^{-1} Hz. The dip in the response curve at 1 Hz is due to the fact that the center of gravity of the pier is slightly below the elastic center of the isolator diaphragms. The rotational transmissibility characteristics below 10^{-1} Hz are a function of the time constant of the level sensor assemblies, the gain of the tilt control circuit, and the time constants of the transmission lines and isolators. Thus, if the translational isolation characteristics of a higher resonant frequency isolation system such as 1 Hz could be tolerated, and the volume in the flapper-nozzle amplifier circuits was halved, the rotational isolation characteristics below 10^{-1} Hz would improve by approximately a factor of 3.

Finally, consider the case of an eccentric moment of magnitude rW_v rotating about the centerline of the pier with a frequency f_r as illustrated in Figure 15. This simulates the condition of an air-bearing turntable with an eccentric load undergoing azimuth rotation. Letting φ_o represent the amplitude of pier tilt that would occur if the tilt control system was not employed, then a plot of $|\varphi/\varphi_o|$ versus frequency would be similar in format to Figure 14. For frequencies f_r less than one-tenth the pitch and roll rotational frequencies of the passive isolation system, φ_o as expressed in seconds of arc is given by

$$\varphi_o = \frac{8.2 \times 10^5 \, rW_v}{a^2 k_z} \qquad (15)$$

where rW_v is in inch-pounds, "a" is in inches, and k_z is the vertical stiffness of the isolation system. For higher values of f_r, φ_o is a function of the dimensions b and c and isolator damping.

In this application, the upper limit of azimuth rotation speeds is 10^{-2} Hz, "a" is 72 inches, and k_z is 1300 lbs/in. Since the pier tilt control system can only compensate for 97 percent of φ_o at this frequency, then based on an allowable residual amplitude of 0.05 seconds of arc for φ_o, the maximum allowable value of rW_v is given by

$$rW_v = \frac{(.05)(72)^2 1300}{(.03)8.2 \times 10^5} \text{ in. lbs.}$$

$$rW_v = 13.6 \text{ in. lbs.}$$

The allowable value of rW_v increases rapidly for lower azimuth rotation speeds. However, increasing the azimuth rotation speeds of the turntable above 10^{-2} Hz (i.e., 1000 earth rate) will require more accurate balancing.

IV. Installation Requirements

The accuracy of the level sensor is effected by the cleanliness of the supply air. Since the pendulum displacement along the centerline of the nozzles is only 3.5×10^{-6} inches for each 0.1 arc second tilt of the pier, dirt buildup on one of the baffles will cause the mean gap to decrease and, therefore, change the null level position of the pier. Through experimentation, it has been determined that the air supply must be oil free and have a dewpoint less than -40° F with a maximum particle impurity of one micron. The supply pressure required for the HCLW system is 70 psig with a short term flow capacity of 20 CFM to compensate for high moments (i.e., 500-foot pounds and up) in the time periods described earlier. The steady-state flow requirements of the system are approximately 5 CFM.

The second parameter of importance is the fluctuations in the environmental temperature. The room temperature in the HCLW installation is maintained at a mean value within $\pm 1/4^{\circ}$F. Under these conditions, no temperature sensitivity of the control system could be detected. However, tests conducted in an uncontrolled temperature environment prior to delivery of the system indicated a temperature sensitivity of 1/4 second of arc per degree Fahrenheit. This error appears to be primarily a function of how accurately the pendulum is initially balanced between the nozzles. Thus, with improved calibration procedures in future systems, the temperature sensitivity should be reduced.

Since the working surface of the pier is a large area (8100 in.^{2}) and since it is suspended by a vibration isolation system having a low vertical natural frequency (0.6 Hz), the pier is susceptible to low frequency pressure disturbances. For example, rapidly opening the door in a room pressurized to 0.02 inches of water gage will cause a momentary tilt of 1 second of arc which decays within a few seconds following door closure. The solution to this problem is the use of an air-lock entrance.

V. Future Capabilities

As man probes deeper into space, the increasing need for improved angular stability and vibration isolation of the piers used to support inertial navigation test equipment has become more obvious. In the near future, stability goals of +0.02 seconds of arc over indefinite time periods will be required. This means that actuators on 100 inch centers must be able to induce motions of less than 5 micro inches. Calculations in Section 3 have shown that a pendulous tilt transducer utilizing air gaging techniques offers the capability of attaining these future goals through the reduction of pivot friction. The flexible diaphragms presently used in the isolators have virtually infinite resolution. Also, the present pneumatic isolation and tilt stability system can be improved in response speed by separating the two functions. The system would consist of a low frequency passive isolation system upon which would be placed a low volume, rapid response, pneumatic or hydraulic actuator that would be servo-controlled by the tilt sensing system. With regard

to azimuth correction relative to inertial space, work is yet to be performed before a suitable reference that is continuously monitorable can be obtained.

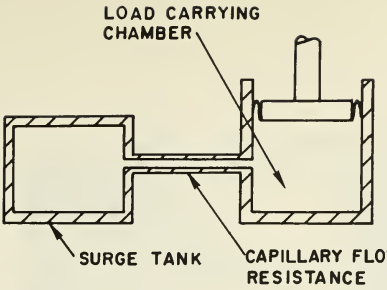
References

1. Weiher, T. E., "Progress in Test Pad Stability," AIAA/ION Guidance and Control Conference, 1965.

2. Kunica, S., "Protecting Laboratory Equipment from Vibration Environments," AIAA/JACC Guidance and Control Conference, 1966.

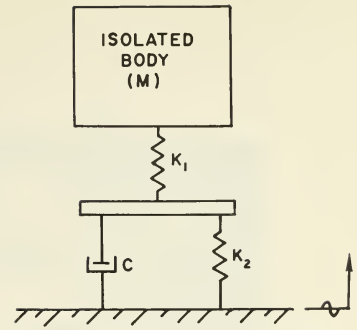
3. Ruzicka, J. E., "Resonance Characteristics of Unidirectional Viscous and Coulomb-Damped Vibration Isolation Systems," ASME Paper No. 67-Vibr-21, 1967.

4. Harris, C. M., and Crede, C. E., Shock and Vibration Handbook, McGraw-Hill Book Company, Inc., 1961.



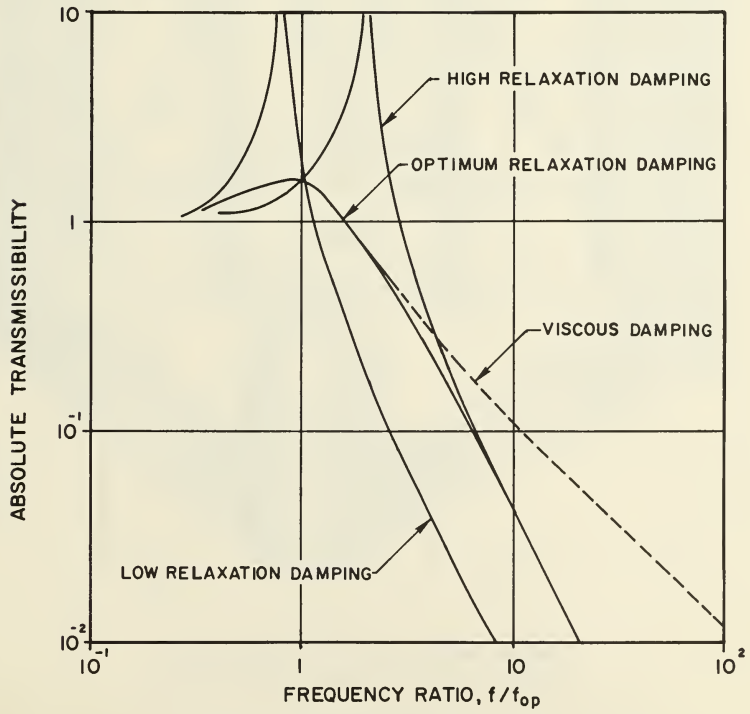
SURGE TANK DAMPING OF A PNEUMATIC SPRING

FIGURE 1



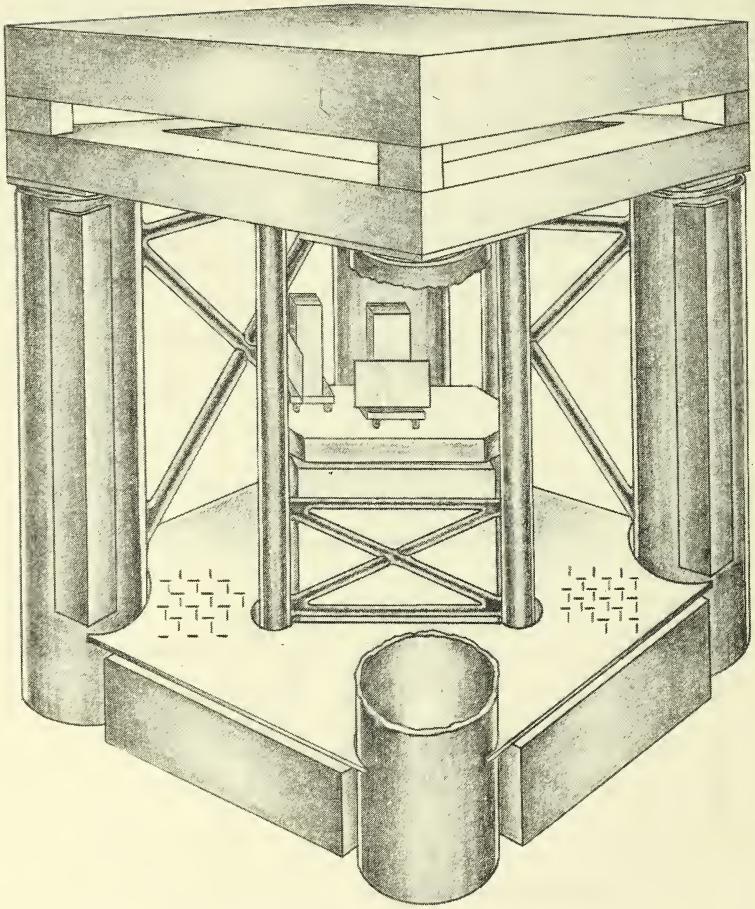
MODEL OF RELAXATION DAMPED ISOLATION SYSTEM

FIGURE 2



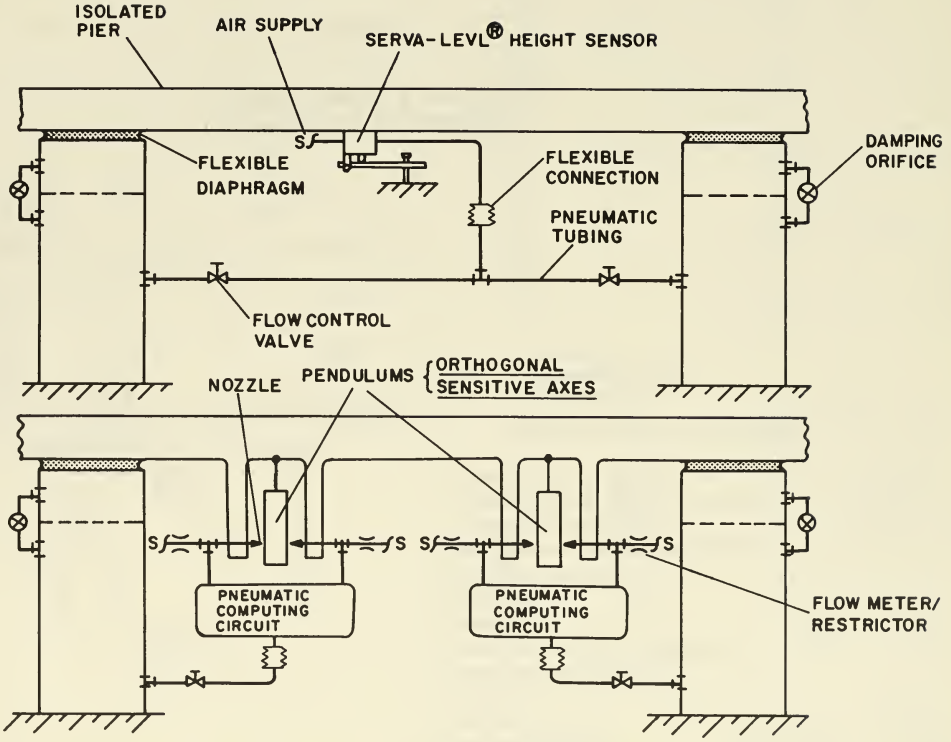
TYPICAL ABSOLUTE TRANSMISSIBILITY FOR RELAXATION AND VISCOUS DAMPED ISOLATION SYSTEMS

FIGURE 3



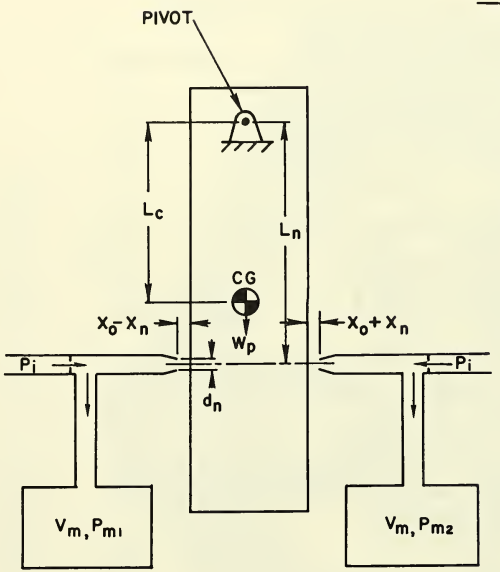
BARRY SERVA - LEVL® STABLE PIER
AT USAF 2802D IG & C LABORATORY

FIGURE 4



SCHEMATIC REPRESENTATION OF STABLE PIER CONTROL SYSTEM

FIGURE 5



The notations ore defined os follows:

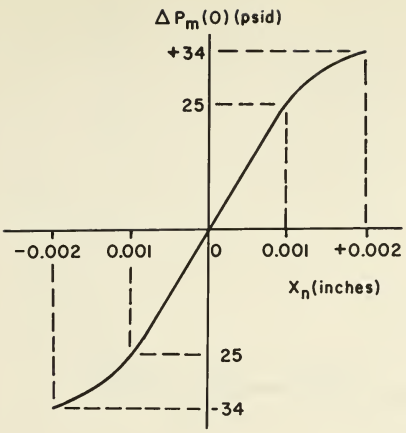
A - orifice area, in^2
d - diometer, in
k - C_p/C_v
P - obsolute pressure, psio
R - gos constant, 2.5×10^5 $in^2/sec^2 \cdot °R$
T - temperoture, °R
V - volume, in^3
W_p - weight of pendulum
X_o - steodystate position of pendulum relotive to nozzle, in
X_n - relotive displacement of pendulum from meon position, in

Subscripts denote the following:

i - oir supply
m - sensing circuit
n - nozzle

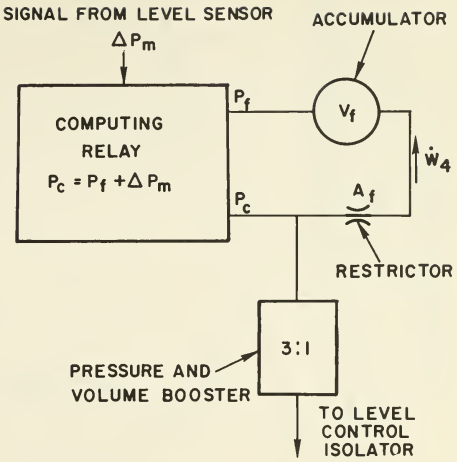
SCHEMATIC DIAGRAM OF
FLAPPER - NOZZLE AMPLIFIERS

FIGURE 6



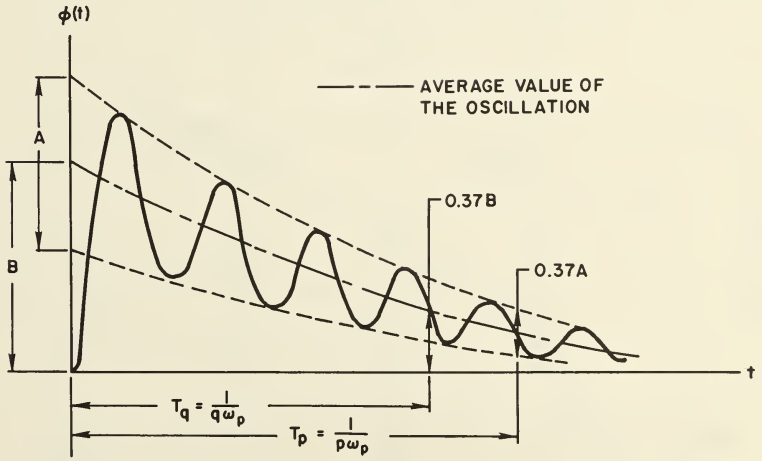
EXPERIMENTAL STATIC GAIN OF FLAPPER-NOZZLE AMPLIFIERS

FIGURE 7



SCHEMATIC REPRESENTATION OF COMPUTING CIRCUIT

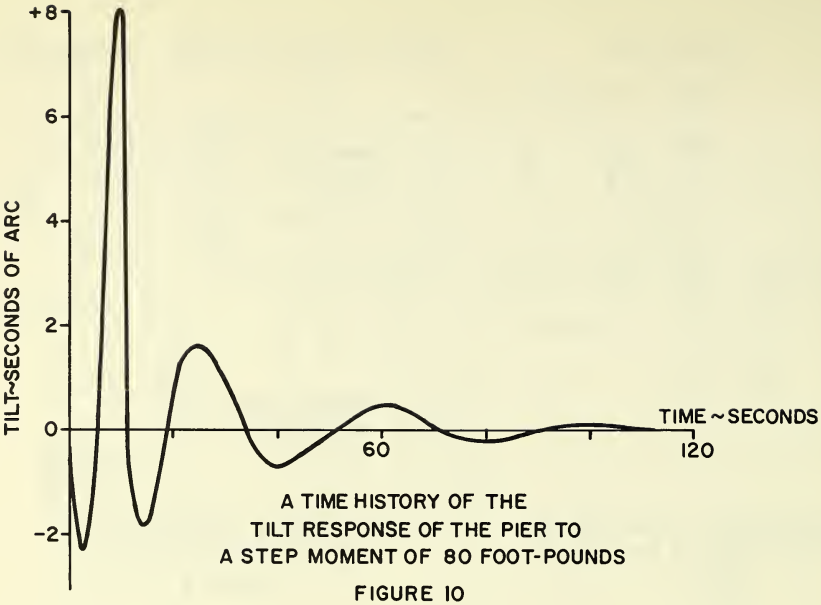
FIGURE 8



TYPICAL ROTATIONAL TIME HISTORY FOR AN ACTIVE TILT CONTROL SYSTEM SUBJECT TO A STEP INPUT

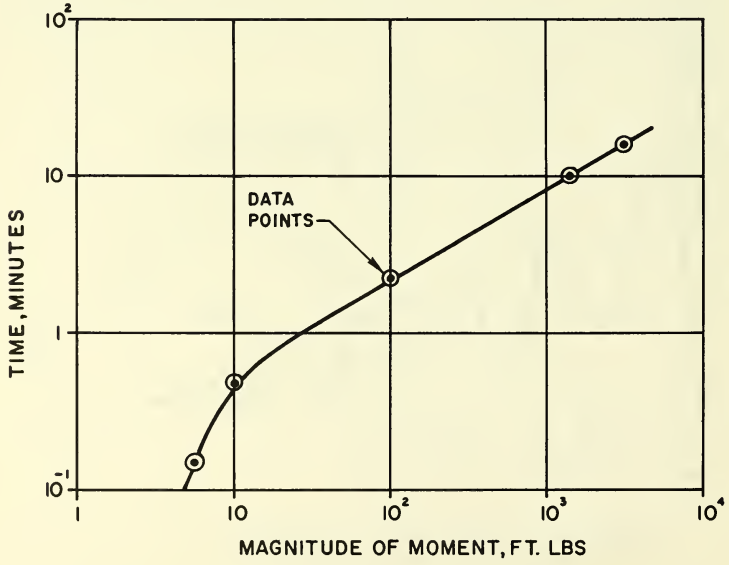
FIGURE 9

67-548



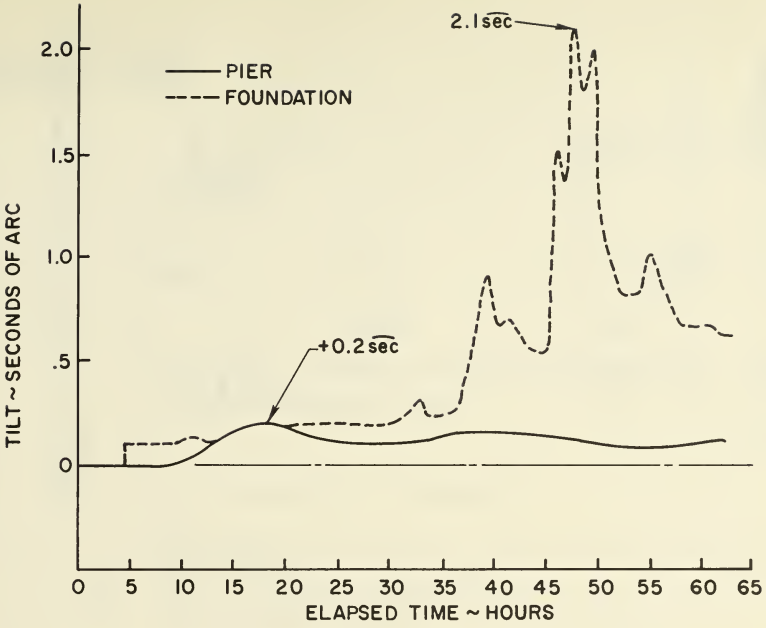
A TIME HISTORY OF THE
TILT RESPONSE OF THE PIER TO
A STEP MOMENT OF 80 FOOT-POUNDS

FIGURE 10



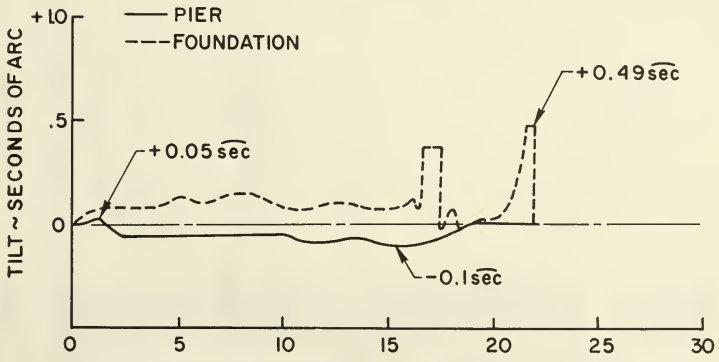
TIME FOR PIER TO TILT STABILIZE
FOLLOWING VARIOUS STEP MOMENTS

FIGURE 11



A TIME HISTORY OF THE PIER TILT POSITION
AS COMPARED TO FOUNDATION TILT

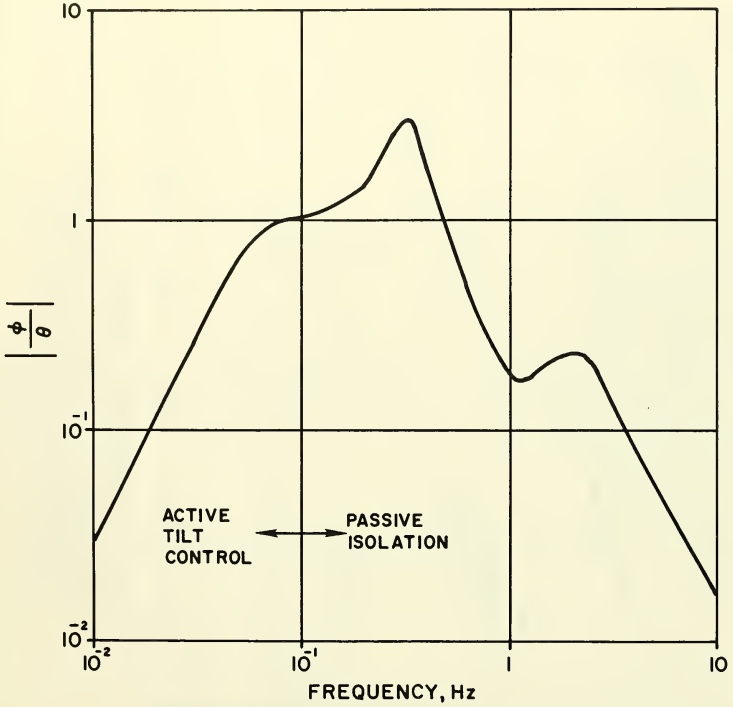
FIGURE 12



A TIME HISTORY OF THE PIER TILT POSITION
AS COMPARED TO FOUNDATION TILT

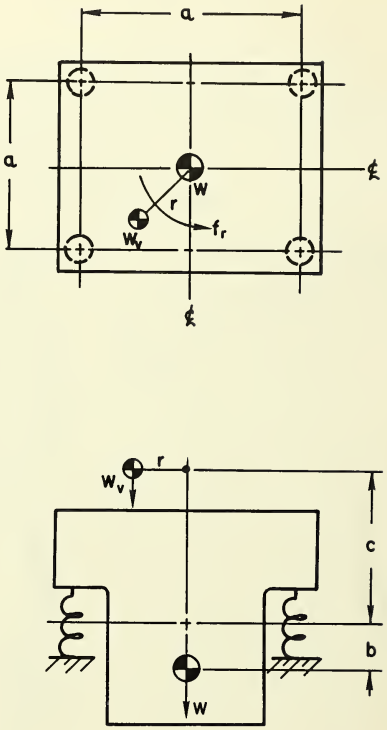
FIGURE 13

67-548



COMPUTED ROTATIONAL TRANSMISSIBILITY OF PIER

FIGURE 14



SCHEMATIC OF PASSIVE ISOLATION SYSTEM SUBJECT TO AN ECCENTRIC ROTATING MOMENT

FIGURE 15

No. 67-549

OPTIMIZED DESIGN OF GROUND-FOUNDED MICROPRECISION SLABS

by

B. E. MARGASON, J. A. BARNEICH,
and
F. M. BABCOCK
Woodward-Clyde-Sherard & Associates
Oakland, California

AIAA Paper
No. 67-549

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

10.05, 10.10

OPTIMIZED DESIGN OF GROUND-FOUNDED MICROPRECISION SLABS

B. E. Margason, Senior Projects Manager
J. A. Barneich, Project Engineer
F. M. Babcock, Senior Projects Engineer

Woodward-Clyde-Sherard & Associates
Oakland, California

Abstract

Increasingly stringent isolation requirements for microprecision slabs, combined with the need to use convenient but often very noisy slab sites, requires not only properly designed solid slabs but also special techniques such as subsurface energy barriers and hollowed slabs. The state-of-the-art in energy-barrier technology is described in detail for both cutoff trenches and sheet pile walls. Subsurface cutoff barriers are shown to be effective only if they are rigid and extend to a depth of at least one-third the Rayleigh wavelength of the lowest frequency of incoming vibrations. Because of wave dispersion, barriers are effective only when placed close to either the vibration source or receiver, or both. In certain situations where external slab dimensions are fixed and the frequency of disturbing vibrations is low, say below 5 to 10 cps, the slab frequency and modal damping can be increased with a corresponding response reduction only by designing a hollow test slab. Parameter studies and a typical case history describe the interplay of these optimizing techniques and demonstrate that only careful application of these concepts, combined with sound engineering judgment and experience, will yield optimum isolation.

I. Introduction

Increased isolation requirements for today's microprecision slabs have necessitated some bold advances in the design of ground-founded slabs. At the present time, microprecision isolation criteria can be met only if slab designs are optimized by using all of the beneficial attenuation techniques available to the designer. This paper describes the methods used for isolating ground-founded slabs from low-level, reasonably steady-state disturbances originating away from the slab. A companion paper[4] deals with slab response and attenuation for high level transient inputs applied directly to the top of a slab.

II. Behavior of Ground-Founded Slabs

The basic behavior of dynamically loaded foundations, including the background theory, has been explained in detail in Reference 1. This discussion will review only the more important of those concepts, and will present new work recently available.

When a foundation moves, it propagates both body and surface waves in the supporting earth mass. These are shown in Figure 1. Knowledge of the body waves is important for two reasons: 1) they can cause destructive resonances if they reflect off of lower layers (that effect will not be considered further in this paper); and 2) they disperse in three dimensions throughout the earth mass. As explained in Reference 1, this dispersion leads to a mathematical damping term, called spatial damping, which will be discussed further below. Knowledge of the surface waves is also important for two reasons: 1) surface waves disperse or attenuate in two dimensions[1]; and 2) they are low-frequency, high-amplitude Rayleigh waves which are particularly disturbing to isolated foundations.

Recent theoretical and experimental work[2,3] has greatly simplified the parameters of foundation vibrations. The equation governing the motions of foundations is of the form,

$$M\ddot{x} + C\dot{x} + Kx = F(t) \tag{1}$$

where x is the motion coordinate; the dots indicate time differentiation; M, C, and K are mass, damping, and stiffness parameters for the foundation; and F(t) is the forcing function. Parameter M is simply the mass of the foundation block. Parameters C and K are, strictly speaking, frequency-dependent variables. For vertical translation, however, Lysmer and Richart[5] have shown that C and K can be taken as constants in a particularly simple form:

$$C = \frac{3.4}{1-\nu} R^2 \sqrt{\frac{G\gamma}{g}} \tag{2}$$

$$K = \frac{4GR}{1-\nu} \tag{3}$$

where R is the radius of the foundation, G is the soil shear modulus, ν is the soil Poisson's ratio, γ is the soil unit weight, and g is the acceleration of gravity. The ratio between the actual damping and the critical damping is a very useful parameter in practical problems. It is shown in a companion paper[4] that for the vertical translation mode, this damping ratio, ζ, becomes simply,

$$\zeta = 0.45\sqrt{\frac{R}{H}} \tag{4}$$

where R is the radius and H is the height of a solid-concrete circular foundation.

For the case of constant coefficients, Equation 1 can be modeled by mass-dashpot-spring systems. The choice of model depends on the problem being solved, as shown in Figure 2. For a response problem, when F(t) acts on the mass, and the motions of the mass are to be limited, Figure 2a is the appropriate model. For isolation problems, when F(t) acts from the ground but the motions of the mass are to be limited, Figure 2b is the appropriate model. For F(t) harmonic, Equation 1 is well known, and the solutions for response and isolation are given respectively in Figures 3a and 3b. For F(t) transient, solutions are described in Reference 4.

The damping due to wave dispersion is shown[5] to depend only on the size and mass of the foundation

block, expressed as a mass ratio[4]. The amount of damping, however, is different for different modes of vibration; Figure 4 shows the dependence on mode. It is important to note that the translational modes have high damping; but the rotational modes have low damping. These dampings are quite important for practical problems. Note, for example, the highly beneficial effect of damping for response problems, Figure 3a. Because of their high damping, translation modes are usually not much trouble for response designs. Rotational modes, however, are notoriously bad for response designs, because they have low damping. In fact, experience shows that the rocking mode is the biggest problem in a response design of a foundation, especially if that foundation is tall and narrow.

For isolation problems, however, damping may be a benefit or a detriment, as is shown by Figure 3b. For foundations designed to have a high natural frequency ($f/f_n < \sqrt{2}$, Figure 3b), damping is beneficial. For foundations with a low natural frequency, however, damping is theoretically detrimental (but, as a practical matter, damping is of little effect in this case). Thus, for high-frequency foundations, rotational modes (low damping) are usually magnified at most frequencies. For low-frequency foundations, however, all modes are attenuated no matter what the damping.

For problems in test-pad stability, isolation is the more important of the two response models. Note from Figure 3b that for order-of-magnitude isolation, the frequency ratio f/f_n must be high. This means that either the input disturbance must have a high frequency, f, or the response frequency, f_n, of the slab must be very low. Soils are very efficient high-frequency filters, so that high frequencies simply are not transmitted very far from the source of a disturbance. Furthermore, because of the conversion of body waves to surface waves, some of the high-frequency content of the source is coupled into low-frequency Rayleigh waves. As a practical matter, then, isolation is obtained in most cases by designing a low-frequency slab; but the slab should also have some damping so that it will not respond to vagrant high frequencies which might occur.

For typical isolation designs, a slab might have a response frequency between 5 and 10 cps, and inputs might lie between 20 and 30 cps. Thus, one often encounters frequency ratios between 2 and 6. For reasonable damping, Figure 3b shows the transmissibility to lie between perhaps 0.5 and 0.05 for these typical cases. Thus a simple slab designed on the basis of transmissibility usually does not isolate by more than about one order of magnitude. This amount of isolation is adequate for many installations such as forge hammers and manufacturing processes. For microprecision test pads, however, more isolation is usually required. For this additional isolation, certain optimizing procedures are required. These will be explained in the next section.

III. Methods of Optimizing Slab Performance

There are essentially three design features available to optimize the performance of an isolation slab. The three are: 1) trenches; 2) sheet pile walls; and 3) hollowed slabs. Each does its job in a different way, and with different results. The three methods will now be discussed.

Subsurface Trenches

Trenches work principally by cutting off the low-frequency, high-amplitude surface Rayleigh waves. Figure 5 is a plot, from the theory of Rayleigh waves, of the vertical and horizontal amplitudes as a function of depth. The depth in that figure is given in terms of wavelength, λ, which may be computed from:

$$\lambda = \frac{C_r}{f_r} \tag{5}$$

where C_r is the wave velocity and f_r is the frequency of the Rayleigh wave. For reasonable soil materials, C_r lies between 200 and 800 fps; and at distances from the source, the frequencies f_r are typically 10 to 30 cps. Thus, wavelengths can lie between perhaps 7 and 80 feet, with typical values usually around 30 feet. Note in Figure 5 that for depths of from 1/4 to 1/2 the wavelength, the horizontal motions are substantially reduced, and the vertical motions are reduced somewhat.

Experience to date shows that the theoretical solution of Figure 5 is conservative for real soils and real foundations. Thus, trenches on the order of 7 to 15 feet deep typically reduce the motions by at least, and usually more than, the theoretical values in Figure 5. In fact, experience shows that any trenching does some good, regardless of the amplitude of the incoming vibrations.

Detailed quantitative studies on the efficacy of trenching are definitely lacking. From what is known of the theory of wave formation and propagation, one would deduce that trenches would be most effective at the receiver slab rather than at the source slab. This is because the Rayleigh waves are well formed away from the source. The limited experience to date, however, indicates that trenches are extremely effective in reducing transmissions from source slabs. Experience shows[6] that trenches are almost totally ineffective below the water table. The reasons for this are not immediately clear: the Rayleigh wave is principally a shearing motion, but water cannot transmit shear. These two anomalous examples show the need for fundamental studies on the topic of trenches.

It is known, however, that the value of a trench depends on the depth dimension of the trench and the distance from the source, both dimensions being fractions of wavelength. For a given material, Equation 5 demonstrates that wavelength is inversely proportional to frequency. Thus for low frequencies (long wavelengths), dimensions must be long; and for high frequencies (short wavelengths), dimensions may be short. For example, Figure 6 shows results of an experiment from Reference 6. The trench, which is 12 feet deep, is located some 7-1/2 feet from the source, which is a variable-frequency sine-wave generator mounted on a concrete slab. The vibrator is operated at various frequencies but constant amplitude, and the motions are measured at the ground surface at distances from the trench. Figure 6 shows that at frequencies higher than about 15 cps the trench is indeed an effective filter. If the trench were deeper, it would be more effective in isolating the lower frequencies. This effect is

shown in Figure 7, where measurements were made as the trench was being dug. Note that, for a vibrator frequency of 12.2 cps, the trench was quite beneficial; however, the benefits depend on the trench depth. For depths of greater than 5 or 6 feet, the cutoff effect did not increase for this soil and this driving frequency. For this case, a loess soil, the dimension of 5 or 6 feet is probably about one-half a wavelength at 12.2 cps.

Sheet Pile Walls

The action of sheet pile walls, although beneficial, is even less understood than the action of trenches. In fact, one could argue convincingly on the basis of wave-propagation theory, that such walls should have only a small effect as a wave barrier. Nevertheless, consulting experiences and data cited by Barkan[6] show that reasonably rigid sheet pile walls act as filters against high-frequency components of earth waves.

Figure 8 shows the results of measurements of the effects of a sheet pile wall. The sheet pile wall is straight and some 30 feet long. It is symmetrical about a centerline through the source, which is a vibrator on a concrete slab. Measurements were taken before and after installation of the wall, with the results shown in the figure. For this case, a loess soil and a vibrator frequency of 16.7 cps, the wall reduced the motions by about one-half. Estimates of the efficacy of a wall must consider the wavelengths involved and the geometry of the wall, as will now be explained.

The effect of wavelength (or frequency) on the efficiency of a sheet pile wall is quite pronounced. Figure 9 shows the effect for vibrator frequencies varied from 10.0 to 31.5 cps (more data are presented in Reference 6). The effect of wall geometry is also pronounced[6]. Figure 10 shows the motions before and after installation of sheet pile walls, for a straight wall, and an enclosing wall. The soil conditions are similar, and the driving frequencies are essentially the same. The benefits of an enclosing wall are obvious.

The soil type as determined by the Rayleigh wave velocity, or shear modulus, also controls the depth of the sheet pile system. Figure 11 shows a plot of effective sheet pile length against velocity and shear modulus; it is important to note that as the soil becomes more competent, i.e., higher velocity or modulus, the sheet piles must extend deeper to isolate against any given frequency. The paradox thus arises that the more competent the soil the more difficult it is to drive the sheet piles to an adequate depth; thus there is a limit to the use of sheet piles in dense soils, especially gravels and cobbles. Figure 11 also demonstrates that the higher the incoming frequency the shorter the required sheet pile depth for effective isolation. Thus, sheet piles isolate high frequencies best.

Although the action of walls and trenches is not yet well understood, it is felt[6] that the action is similar to a diffraction process. That is, the benefits are greatest next to the wall or trench; and at some distance away, the benefits are trivial.

This section demonstrates that the use of trenches and sheet pile walls can lead to transmissibility reductions of from 0.8 to 0.1, and for a rule-of-thumb, attenuations of about 0.5 can be reliably attained for most amplitudes. Walls and trenches may be designed only by considering frequencies, soil profile and properties, wall strength and geometry, and depth of barrier. The greatest attenuation benefit might be one order of magnitude (high frequency, weak soil, small slab, and enclosing wall); and barriers would not be used if the computed benefits were less than 10 or 20 percent.

Further isolation can often be obtained by balancing the mass and geometry of the slab itself. This requires considering the response characteristics of hollow slabs, which is discussed next.

Hollow Slabs

Situations arise where the damped, natural frequency of a proposed microprecision slab, even when shielded, is closely matched with ambient frequencies; this results in large slab response. Yet, it is often impossible to change the external dimensions (radius and height) of the slab to effect a change in response. Therefore, only the mass of the slab can be changed, and this is done by hollowing out the slab.

It can be seen from Equations 2 and 3 that the terms depend only on foundation geometry, R. However, in deriving Equation 4, the basic damping relationship of:

$$\zeta = \frac{C}{2\sqrt{KM}} \qquad (6)$$

was used. Equation 6 does depend on mass, therefore both mass and geometry must be related to obtain a useful value of damping ratio, ζ. This is done by defining the mass ratio b as:

$$b = \frac{Mg}{\gamma R^3} \quad \text{(for translation modes)} \qquad (7)$$

and

$$b = \frac{Ig}{\gamma R^5} \quad \text{(for rotation modes)} \qquad (8)$$

As Figure 4 indicates, the damping in various modes is dependent on mass ratio.

The interplay of mass and geometric parameters can best be demonstrated by considering a slab which is a right circular cylinder having a radius to height ratio, R/H, of two. The slab is hollowed out by various amounts and the change in its frequency, f_n, is given for both the vertical and rocking modes in Figure 12.

As Figure 12 indicates, the natural frequency of the slab increases as the slab is hollowed out; however, the change is greater in the rocking mode. For instance, when the slab mass is reduced by 50 percent, the rocking frequency increases 23 percent, while the vertical frequency increases by only 12 percent. A similar relationship is shown in Figure 13 between damping and mass reduction for both modes. This plot demonstrates that damping increases as the slab is hollowed; however, the rate of increase is even greater than for frequency. The interplay of these effects is used to reduce overall slab response.

By applying the findings of Figures 12 and 13 to the transmissibility curve of Figure 3, one can

decide if a hollow slab design will result in isolation. The only instance when a hollow slab is desirable is when the frequency ratio, f/f_n, prior to hollowing is less than one; hollowing further decreases f/f_n hence reducing transmissibility. Similarly, damping increases by hollowing are beneficial only if f/f_n is initially less than $\sqrt{2}$. Similar conclusions may be applied to response problems, Figure 3a, except that a damping increase is beneficial at all frequency ratios.

At first glance, it may seem that hollowing has little value if incoming frequencies, f, are very high; this is true, unless hollowing is used in combination with sheet piles, which will block out high frequencies. Thus, the hollow slab concept nicely applies to the lower frequencies which ordinarily bypass the barriers, and the lower frequencies are the greatest problem when designing microprecision slabs. The hollow slab concept is of greatest value in tight clearance situations where sheet piles are permissible.

IV. Case History
Brooklyn Microprecision Slab

The inter-relations and uses of the optimization techniques discussed in this paper can be demonstrated by a recent case history. In 1966, a large microprecision calibration slab for ship gyros was proposed for a site in Brooklyn, New York; alternate slab locations were not acceptable. This slab was to be nominally 30 feet in diameter, and the desired response specifications were to be smooth curves faired between points as follows (see Figure 14):

Type of Response	Criteria response ratio	frequency
constant force	1.0	2 cps
	0.2	10 cps
	transmissibility	**frequency**
transmissibility	1.0	2 cps
	0.2	10 cps

Maximum ambient vibrations of the surface soils at the site were as follows:

Mode	Frequency	Amplitude	Acceleration
vertical	9-10 cps	600μ inch	6100μ g
horizontal	5- 6 cps	4-600μ inch	1500μ g

These inputs were caused by truck traffic about 200 feet away and by machine shop and air conditioner machinery in an adjacent building.

The soils at the site were very marginal for a stable slab location; they consisted of about 20 feet of loose to medium dense cinder and sand fill with a water table at 11 feet below grade; the shear modulus of the fill was 5000 psi. Below the fill a layer of medium stiff clayey peat about 6.5 feet thick was encountered with a shear modulus of 1500 psi. Beneath the peat and extending at least 50 feet deep, was a medium-dense, clayey sand with occasional clay seams and boulders; this soil had a shear modulus of at least 10,000 psi.

For several reasons, it became apparent that the desired size of the facility combined with the ambient vibration levels and the water table and peat locations, limited the external slab geometry to about 8 feet thick by 30 feet in diameter. Utilizing the techniques discussed earlier in this paper, it was estimated that a solid slab would have natural frequencies and damping in the various modes as follows:

Mode	Damping %	Response Frequency, cps
vertical	70	2 to 3
horizontal	45	3 to 4
rocking	20	7 to 8
torsion	10	7 to 8

Based on these results it became clear that the solid slab would not even approach the response criteria desired. It was therefore decided to hollow out the slab, while preserving its external geometry, thus altering the damping and frequencies to the following values:

Mode	Damping %	Response Frequency, cps
vertical	90	3 to 4
horizontal	60	4 to 5
rocking	25	9 to 10
torsion	10	7 to 8

The above results represented the optimum configuration which could be achieved by mass and geometric changes alone, and typical response curves for the vertical and rocking modes were drawn, see Figures 14 and 15.

Obviously, all response criteria were still not met so further optimization was obtained by using a trench and a continuous steel sheet pile barrier around the slab. In order to design the barrier it was necessary to recognize several facts: the lowest incoming ambient frequency was about 5 cps; the Rayleigh wave velocity of the fill was about 480 fps; the soil profile of surface fill over low density peat suggested that most of the 5 cps ambients would travel as surface waves in the fill (a wave guide effect with the peat acting as a natural energy attenuator). Using Equation 5 and the screening considerations discussed in Section III, it was concluded that λ was about 100 feet, hence a 35-foot-deep closed cutoff wall would reduce vibrations above 5 cps by 50 percent; reductions of as much as 75 percent could result since the slab diameter was only 0.3 λ, and an 8-foot trench was needed anyway for construction reasons.

This reduction allowed a redrawing of the response curves of Figures 14 and 15 by reducing all response ratios at and above 5 cps by 50 to 75 percent. This operation is demonstrated for transmissibility in the rocking mode, see Figure 16. A smooth fairing between the unshielded curves and the shielded curves was done in the vicinity of 5 cps, assuming an orderly development of the diffracted or screened zone at these input frequencies. Using this technique, the criteria for the vertical mode were met. However, it can be seen from Figure 16 that the desired response was still not completely obtained in the rocking mode. After consultation with the client, it was felt that in view of the economics involved, the nature of the site and the use of conservative ambient-vibration assumptions, the shielded hollow-slab design was acceptable.

The resulting design is given in Figure 17, which shows both slab and sheet pile details. For

ease of forming and construction, the circular slab was altered to an octagonal shape. Grout pipes were included for possible post-construction mass tuning. A suspended working floor with a very low natural frequency was provided over the slab to isolate it from pedestrian-induced vibrations, yet still prevent air coupling between the floor and slab. This floor technique is common in micro-precision slabs and is similar to the approach used at Martin, Denver[7], in 1964. Slab construction is slated for late 1967, and response measurements are planned both during and after construction. These results will be reported as they become available.

V. Conclusions

This paper describes recent design techniques available for microprecision slabs subjected to steady-state ground-induced inputs. Techniques for handling on-slab transient inputs are described in a companion paper[4].

Isolation of ground-founded slabs may be optimized by initially sizing a solid slab using current soil dynamics techniques to achieve maximum frequency mismatch. Subsurface trenches or sheet pile barriers extending to depths of at least one-third the disturbing wavelengths may then be used to extend the degree of isolation. In some cases where tight clearance problems exist the slab may be hollowed to alter its response; however, hollowing should be done only with full knowledge of the interplay of damping, frequency, and response.

These design techniques must be tempered with experience, judgment, and a full understanding of the site and soil conditions to yield a design which is effective, economical, and avoids construction difficulties.

References

1. McNeill, R. L., B. E. Margason, and F. M. Babcock: "The Role of Soil Dynamics in the Design of Stable Test Pads," Amer. Inst. Aeronautics and Astronautics, Vol. Tech. Papers, Guidance and Control Conference (Minneapolis), New York, 1965.

2. Lysmer, J.: Vertical Motion of Rigid Footings, Ph.D. thesis, University of Michigan, 1965.

3. Lysmer, J., and F. E. Richart, Jr.: "Dynamic Response of Footings to Vertical Loading," Proc. of J. of Soil Mechanics and Foundation Division, ASCE, Vol. 92, SM-1, January 1966.

4. McNeill, R. L., B. E. Margason, and J. A. Barneich: "Design of Test Pads for Transient Loadings," Amer. Inst. Aeronautics and Astronautics, Vol. Tech. Papers, Guidance and Control Conference (Huntsville), New York, 1967.

5. Whitman, R. V.: "Analysis of Foundation Vibrations," Vibrations in Civil Engineering, Butterworths, London, 1966.

6. Barkan, D. D.: Dynamics of Bases and Foundations, McGraw-Hill, 1962.

7. Mathis, L. O., J. R. Stephens, and S. C. Wright: "Design and Construction of an Inertial Test Facility," Amer. Inst. Aeronautics and Astronautics, Vol. Tech. Papers, Guidance and Control Conference (Minneapolis), New York, 1965.

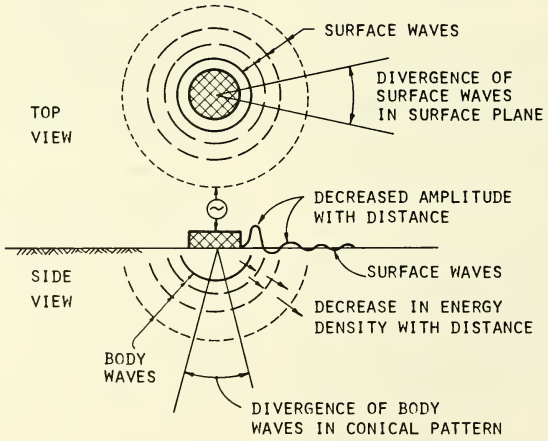
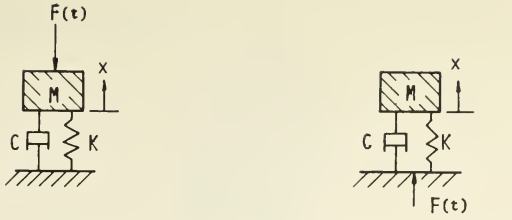
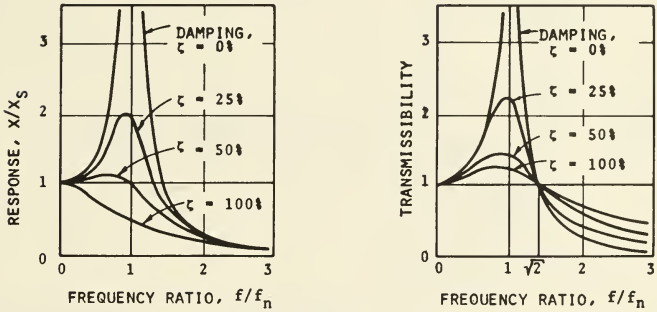


FIG. 1 - FOUNDATION DAMPING



(A) RESPONSE PROBLEM (B) ISOLATION PROBLEM

FIG. 2 - MODELS FOR FOUNDATION DYNAMICS



(A) RESPONSE PROBLEM (B) ISOLATION PROBLEM

FIG. 3 - RESPONSE CURVES

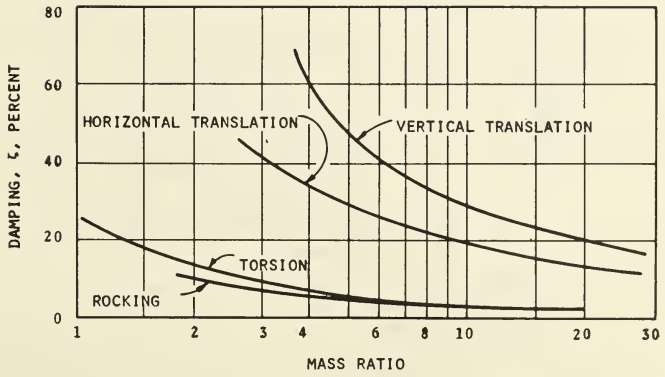


FIG. 4 - MODAL DAMPING RATIOS
(reference 5)

67-549

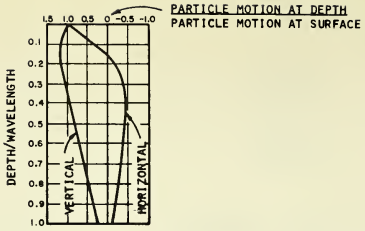


FIG. 5 - DECAY OF RAYLEIGH WAVES WITH DEPTH
(reference 6)

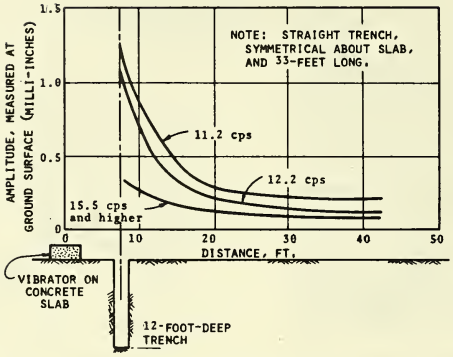


FIG. 6 - SCREENING BY TRENCH
(reference 6)

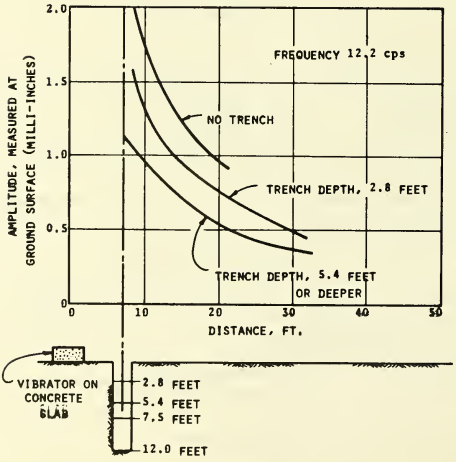


FIG. 7 - EFFECT OF TRENCH DEPTH AT A LOW FREQUENCY
(reference 6)

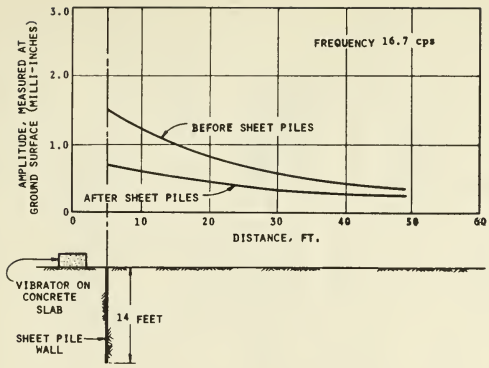


FIG. 8 - SCREENING BY SHEET PILES
(STRAIGHT SHEET PILE WALL)
(reference 6)

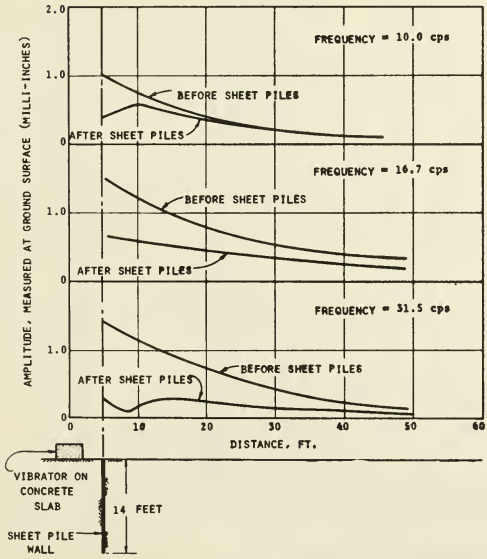


FIG. 9 - EFFECT OF FREQUENCY
(reference 6)

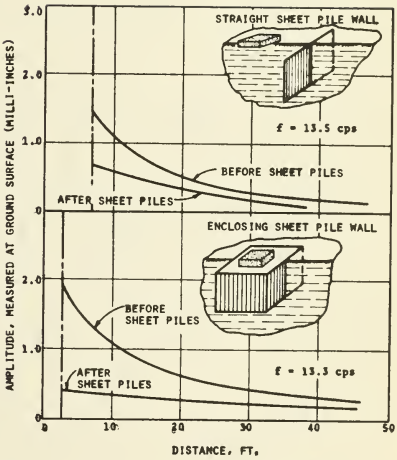


FIG. 10 - EFFECT OF SHEET PILE GEOMETRY
(reference 6)

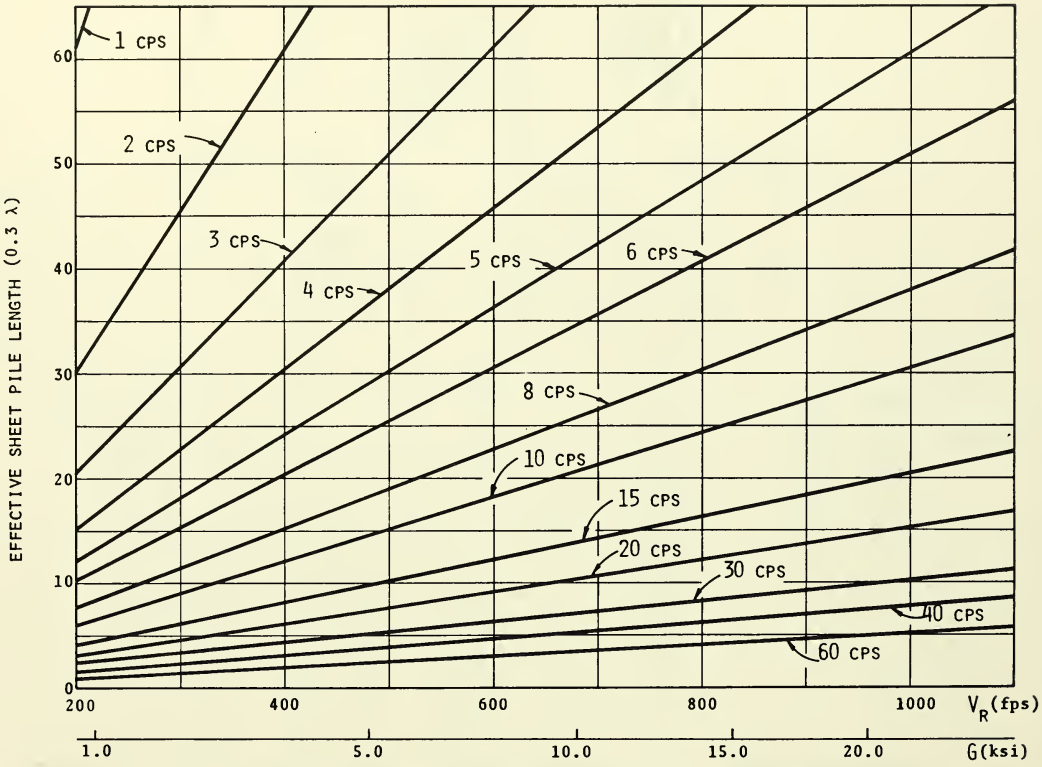


FIG. 11 - REQUIRED SHEET PILE CUTOFF FOR VARIOUS SOILS AND INCOMING FREQUENCIES

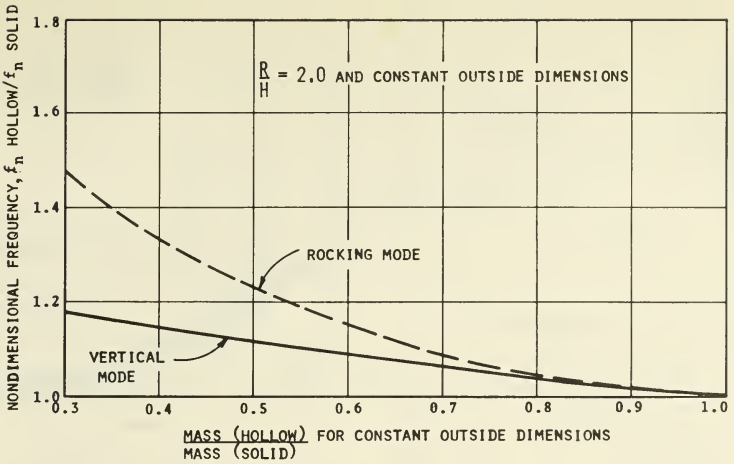


FIG. 12 - EFFECT ON FREQUFNCY WHEN SLAB IS HOLLOWED

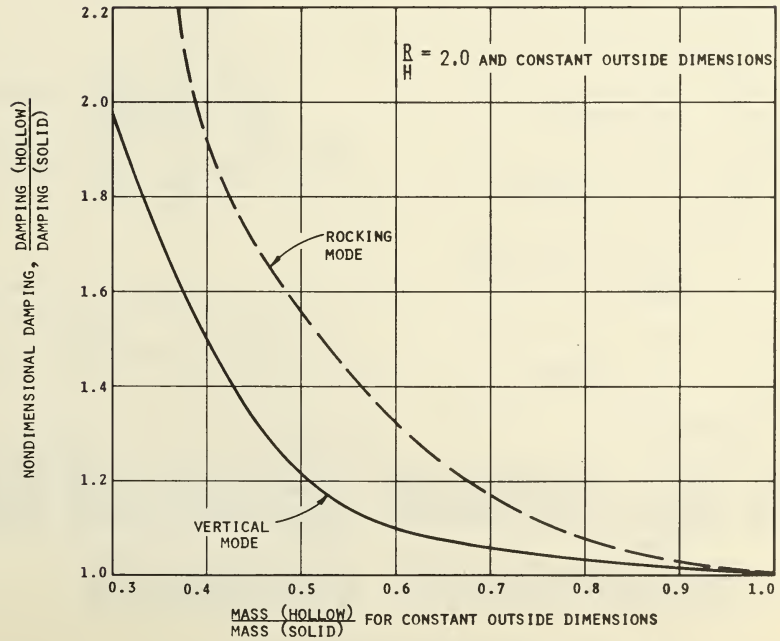


FIG. 13 - EFFECT ON DAMPING WHEN SLAB IS HOLLOWED

67-549

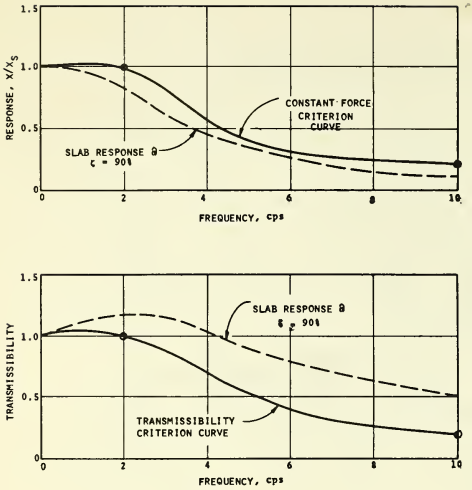


FIG. 14 - RESPONSE CURVES

VERTICAL MODE-UNSHIELDED HOLLOW SLAB (R=15 FEET, H=8 FEET)

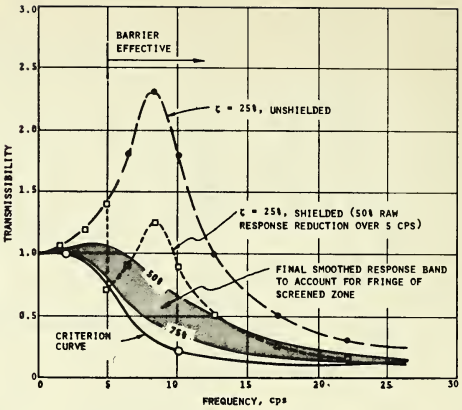


FIG. 16 - RESPONSE CURVES

EFFECT OF SHIELDING WALL AROUND HOLLOW SLAB (R=15 FEET, H=8 FEET)-ROCKING MODE

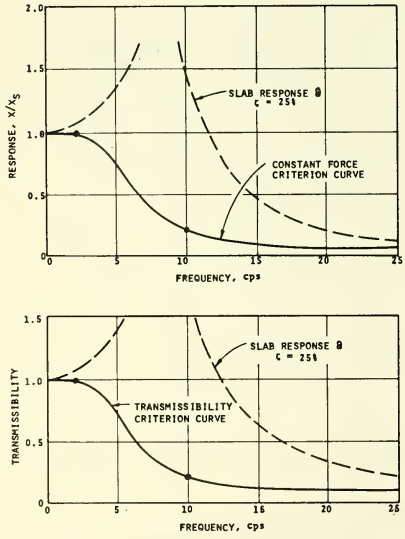


FIG. 15 - RESPONSE CURVES

ROCKING MODE-UNSHIELDED HOLLOW SLAB (R=15 FEET, H=8 FEET)

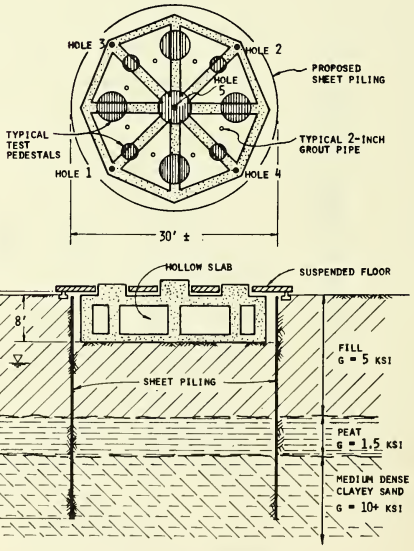


FIG. 17 - DETAILS OF HOLLOW SHIELDED SLAB

No. 67-550

DESIGN OF TEST PADS FOR TRANSIENT LOADINGS

by

ROBERT L. McNEILL, B. EDWARD MARGASON,
and
JOHN A. BARNEICH
Woodward-Clyde-Sherard & Associates
Oakland, California

AIAA Paper
No. 67-550

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

10.05, 10.10

DESIGN OF TEST PADS FOR TRANSIENT LOADINGS

Robert L. McNeill, Director of Special Projects
B. Edward Margason, Senior Projects Manager
John A. Barneich, Project Engineer

Woodward-Clyde-Sherard & Associates
Oakland, California

Abstract

The complexities of transient analysis for test slabs on soil have long been recognized, but until recently such an analysis could be approximated using only steady-state assumptions. The paper demonstrates that it is now possible to make a meaningful engineering analysis of a transiently loaded foundation on soil by the graphical phase-plane method. This procedure is described and is shown to yield accurate results even for extremely complicated transient inputs. An evaluation of the older steady-state approximation shows that it may be very unsafe for cases where the input frequency is greater than the natural slab frequency. In some instances, depending on the input pulse shape and foundation damping, it may even be unsafe when the input frequency is less than the natural slab frequency. In the case where the input and natural frequencies are nearly the same, the steady-state approximation is conservative, especially for low damping. Parametric studies also show that the slab response is strongly dependent on the soil stiffness and foundation damping. The paper concludes with a design case history demonstrating that a low, broad test pad with sufficient structural strength is best for transient loads.

I. Introduction

Until recently, it has been necessary to design all dynamically loaded foundations by harmonic (or steady-state) methods, even though the actual loads may have been transient. By utilizing some recent advances in the theory of foundation vibrations, and by making use of standard methods for the analysis of transient loads on simple systems, safer and more economical foundation designs can be made. The resulting optimized design of a transient-loaded test pad then facilitates a more accurate measurement of the experimental event for which it will be used.

In this paper, the new theoretical advances will be explained and the analytical methods will be reviewed. Then the results of some parametric studies will be presented to show the effects of load function, soil properties, and foundation size. Finally, a design case history will be presented to illustrate the usefulness of the techniques.

It should be noted that this paper is limited to a discussion of the design of test pads subjected to direct transient inputs. The related subject of microprecision slab design and isolation is discussed in a companion paper[11].

II. Dynamic Foundation Motions

The motions of dynamically loaded foundations have received a great deal of attention in the literature of recent years[1,2,3,4]. The mathematics and physics of the problem are solved[5,6] to the extent that research[7] and engineering case histories[8,9] have been presented and interpreted. As has been explained in several of the references, it is necessary to consider the wave propagation aspects of the problem: as the foundation moves, it generates waves which propagate away into the earth mass. Those waves which propagate away represent a loss of energy, or a damping. The reaction of the soil against the foundation represents a stiffness. By proper arrangement of the basic equations of motion[6], it is possible to derive a simple differential equation governing the motions:

$$M\ddot{x} + C\dot{x} + Kx = F(t) \tag{1}$$

where x is a motion coordinate; the dots are time differentiations; M, C, and K are mass, damping, and stiffness parameters; and F(t) is the forcing function. Equation 1 is a well-known form, which has a simple mass-spring-dashpot model, as shown in Figure 1. However, the parameters C and K depend on frequency, so that the simple model is somewhat limited. Within restricted ranges, usually around resonance, response curves for foundations are routinely computed for harmonic problems[8], to help in design, in detuning, and in augmentation.

Until recently, the problem of dynamic foundation analysis was severely restricted by the frequency dependence of the stiffness and damping parameters, K and C. Recently, Lysmer and Richart[3] made detailed studies of the effects of this frequency dependence on the end result, the response of the foundation. Their calculations showed that the computed response was rather insensitive to the frequency dependence of the parameters; and a series of large-scale footing experiments verified the calculated conclusions. For the vertical motions of a circular footing of radius R, the parameters C and K can be taken as constants, for engineering problems, with the values

$$C = \frac{3.4}{1-\nu} R^2 \sqrt{\frac{G\gamma}{g}} \tag{2}$$

$$K = \frac{4\,GR}{1-\nu} \tag{3}$$

where G is the shear modulus for the soil, ν is Poisson's ratio for the soil, γ is the unit weight of the soil, and g is the acceleration of gravity. Usually, for convenience, the damping is expressed as the damping ratio[10], ζ, defined as

$$\zeta = \frac{C}{2\sqrt{KM}} \tag{4}$$

For a candidate foundation design at a given site of known soil properties, the parameters C, K, and M are known so that the damping ratio can easily be calculated.

Now consider, for example, a circular foundation of radius R and height H. Let the foundation be made of solid concrete, so that the ratio of concrete density to soil density is about 1.5. Manipulation of Equations 2, 3, and 4 will yield, for Poisson's ratio of 0.25:

$$\zeta = 0.45\sqrt{\frac{R}{H}} \tag{5}$$

Equation 5 shows that damping ratio depends only on foundation geometry expressed as the ratio of radius to height, R/H, if variations in Poisson's ratio are neglected. Equation 5 is plotted in Figure 2 for ease of reference. Note that for usual design, where R/H lies between 2 and 4, the damping values are high for vertical translation.

Many times, however, foundations are not solid, not circular, or not of constant height. Furthermore, rotational as well as translational modes must be considered in design. For these reasons, it has been found convenient to define the mass ratio, b:

$$b = \frac{M}{\gamma R^3} \tag{6a}$$

$$b = \frac{I}{\gamma R^5} \tag{6b}$$

Equation 6a is the mass ratio for translation modes; and Equation 6b is the mass ratio for rotational modes, where I is the mass moment of inertia about the axis of rotation. Note that the mass ratio combines the foundation geometry and foundation weight. The variation of damping ratio with this defined mass ratio is given in Figure 3 for four modes of motion. The data on Figure 3 lead to two important conclusions: 1) damping is large for translational modes but is small for rotational modes; and 2) damping decreases with increasing mass or decreasing radius for all modes. Thus, if high damping is desirable, as for a transient-loaded foundation, large, low, light foundations are used. The past practice of using arbitrarily heavy foundations thus is not necessarily correct; and that practice can be quite unconservative if the mass is attained by building a high foundation.

These considerations can be quite important when designing for transient loads. In this sort of problem, one usually wants: to minimize the response while the transient load is acting, to minimize the number of oscillations after the load has ceased to act; and to minimize the transmission of energy to nearby instruments and structures. These design features are analyzed by rather simple methods which will now be described.

III. Basis For Transient Load Analysis

In the preceding paragraphs, the parameters, M, C, and K for a foundation have been developed: M is simply the mass of the foundation; C is the damping, most conveniently expressed as damping ratio as in Figure 3; and K is the stiffness given, for vertical translation, by Equation 3. Thus, to engineering accuracy the three parameters may be taken as constants in the differential equation of motion, Equation 1. That differential equation is well known, and extensive solutions have been published[10].

For transient-load designs, there are essentially three methods of solving engineering problems. The first, and simplest, is to convert the transient load to a harmonic load by applying some quarter-wavelength or other arbitrary rule. Although it has been recognized for some years that this harmonic-equivalent method is not accurate, it has been used extensively on the argument that it is conservative for design. Perhaps a more pertinent argument would be that, until recently, only harmonic analyses were available.

The second method is to solve the differential equation of motion, Equation 1, for the given transient loading function F(t). Solutions by this method were not generally possible until about 1965 or 1966, when explicit but approximate equations were derived for the stiffness and damping parameters, for example, Equations 2 and 3. However, because many loading functions do not have simple expressions, solutions to the differential equation can often be made only in series form.

The third, and probably best, method is to make a graphical solution by the phase-plane method. This method is well described in the literature[10], and will be explained only briefly here. The method has proved so useful and simple in recent practice that it is becoming the exclusive method for solution of transient-load foundation problems.

The phase plane is a plot of the locus of points cut by the tip of a rotating vector. As shown for the undamped case in Figure 4a, the axes of the phase plane are velocity (divided by natural frequency) for the abscissa, and displacement for the ordinate. The radius vector rotates at an angular velocity, p, which is the natural frequency of the vibrating system the phase plane represents:

$$p = \sqrt{\frac{K}{M}} \tag{7}$$

A given angle, $\Delta\theta$, on the phase plane represents real time:

$$\Delta t = \frac{\Delta\theta}{p} \tag{8}$$

Projections, against real time, of the velocity and displacement axes thus represent the time-history of the motions, as shown in Figure 4a. Thus, if the system characteristics are known (or assumed for candidate designs) so that p may be computed, then the rate at which the vector rotates is known, and angles on the phase plane can be converted to real time. The radius of the circle is the maximum amplitude, which can be calculated from the load function as will now be explained.

For purposes of graphical solution, any load function can be represented with a series of small square-wave steps. The number of steps is determined by the accuracy required in the particular engineering problem. The force F of any step can be converted to an equivalent static displacement, X_s, by dividing by the stiffness:

$$X_s = \frac{F}{K} \tag{9}$$

For simplicity, the load function can then be shown on the displacement plot, as is done in Figure 4b. The load step acts for a given time T_1; and in this time the phase-plane vector will rotate through an angle,

$$\theta_1 = pT_1 \tag{10}$$

For reasons explained in Reference 10, the origin of the phase plane will be (O, X_S) as shown by point 1 on the phase plane of Figure 4b. The radius, $\overline{10}$ rotates for the time T_1 through the angle θ_1 to trace out the arc $\overset{\frown}{0ab2}$; and each point on that arc represents a displacement-time point, as shown on the displacement-time plot. At the end of time T_1, the radius is in position $\overline{12}$, the load stops acting, and free vibration commences, governed by the radius R, shown by $\overline{02}$, rotating about the origin of the phase plane. Projections of the vector trace out the displacement-time history 3-4-5-6, etc., as shown. In the same way, projections from the $\dot{x}/p$ axis trace out the velocity-time history as shown.

The phase plane is used to solve damped-vibration problems in the same way, except for three modifications. The first modification is that the $\dot{x}/p$ axis is skewed to the abscissa by an angle, β, defined by:

$$\tan \beta = \frac{\zeta}{\sqrt{1-\zeta^2}} \tag{11}$$

The displacement axis remains vertical, and coordinate points are measured parallel to the skewed axes. A set of such damped-vibration axes are shown on the phase plane of Figure 5. The second modification is that the radius vector rotates at the damped frequency, d, given by:

$$d = p\sqrt{1-\zeta^2} \tag{12}$$

The third modification is that the radius vector of the phase plane is continually decreasing in length due to the damping. It is easily shown[10] that if the radius vector is some length r_1, at a given point on the phase plane, then after a rotation $\Delta\theta$ that radius vector will be a shorter length r_2, given by:

$$r_2 = r_1 e^{-\Delta\theta \tan \beta} \tag{13}$$

where e is the Napierian base. Any angle $\Delta\theta$ is still related to real time by Equation 8. The foregoing items are illustrated on the phase plane and the displacement plot of Figure 5a. The operations for plotting the velocity-time history are the same as has been explained, except that the skew of the velocity axis must be taken into account as shown in Figure 5a.

The procedure for plotting forced damped motions is the same as for plotting forced undamped motions, except that the three modifications described above must be taken into account. Figure 5b shows an example of computations of forced, damped motions.

Use of the simplified constants and equations of motion, along with the phase-plane method of graphical analysis has opened wide possibilities for more realistic designs of foundations subject to transient loads. In the next section, some broad parametric comparisons will be made to illustrate the effects of some foundation variables.

IV. Parametric Comparisons

The parameters involved in a transient-load foundation problem fall into four groups:

1. Soil Properties
   - density
   - stiffness properties
   - damping properties
   - homogeneity or layering
   - isotropy
2. Foundation Properties
   - size, plan and elevation
   - shape
   - weight and inertia
3. Input Load
   - point of load application
   - time history of load
4. Criteria
   - what is the desired response of the designed foundation under the given input load?

These parameters will now be discussed and simplified. For the purposes of this paper, it will be assumed that complete failure of the foundation is not a possibility; and that static long-term settlement is of no concern. In a real problem, of course, both of these items would be given close attention in the design.

For most soils, in-situ density is a relatively unimportant variable for dynamic problems. Usual density, or unit weight, values lie between 90 and 120 pcf; but exceptional values can range from 70 to 150 pcf. For an actual design, the measured in-situ density would be used (care should be taken not to use the dry density, which is the value usually measured). For the remainder of this discussion, a constant 100 pcf will be taken for soil density.

For dynamic analyses, stiffness properties are expressed by any two of the elastic parameters: Young's modulus, E, shear modulus, G, and Poisson's ratio, ν. Practice shows that G and ν are most convenient. Soils are nonlinear, so that elastic parameters are, at best, approximations taken for convenience of computation. Usually, the parameters are computed as secants or tangents to a given stress level on a stress-strain curve. It is imperative that the method of computing the elastic constants be given if the design is to be checked. For this discussion, typical values of G and ν will be used, and in one phase of this study the effect of changing G on the response of a given slab subjected to transient inputs will be explored.

Soil damping properties are not well understood, but some broad generalizations can be drawn: the lowest damping values, only a percent or so, are found in dense, dry, cohesionless materials; and the highest damping values, on the order of 10 percent or more, are found in soft, partially saturated, cohesive soils. As Figure 3 shows, the spatial damping for translational modes can easily be greater than 50 percent; but the damping for rotational modes is seldom greater than 10 or 20 percent. For translational modes, soil damping is negligible and is usually neglected for ease of computation. For this discussion, soil damping will be neglected.

Attention must be given, however, to spatial damping as it may have a substantial effect on slab response.

Most soil deposits are nonhomogeneous, layered, and anisotropic. For accurate designs, the layering and degree of anisotropy must be considered. The techniques of accounting for these complicated systems are, however, beyond the scope of this paper; therefore, the discussion to follow will assume homogeneity and isotropy.

For simplicity in this discussion, a solid, right-circular cylinder will be assumed for the foundation. The potentially beneficial effects of using hollow foundations are discussed in Reference 11. Furthermore, the foundation block will be assumed completely rigid with respect to the stiffness of the soil; and the block will be assumed to lie at the ground surface (effects of burial are discussed in Reference 4).

It is an anachronism of present-day foundation dynamics that the foundation can often be designed with more precision than the input load can be determined or specified. For this discussion, only simple load shapes will be used; no criteria will be specified.

For this parametric study, then, only the variables of soil shear modulus, spatial damping, and radius-to-height (R/H) of foundation will be considered. Furthermore, in order to limit the discussion, only a single load function acting in vertical translation will be explicitly studied.

The effect of the load function will be considered first. For purposes of illustration, the simple load function shown in Figure 6a will be used: the load pulses are square, each lasting a time T/3, with the middle pulse twice as large as the first and third pulses. The total positive-phase duration of the load function is the time T. Only the undamped case will be analyzed (results for the damped case are given later). For a given system with mass M and stiffness K, the natural undamped frequency p is given by Equation 7. The loads F and F/2, Figure 6a, are converted to static deflections by dividing each by the stiffness K, according to Equation 9. For an undamped system, the characteristics are completely described by the natural frequency p, which is conveniently related to the natural period, T_n, by

$$p = \frac{2\pi}{T_n} \tag{14}$$

Then for a time increment T/3 of one of the three load pulses, the radius vector in the phase plane will rotate through,

$$\Delta\theta = p\frac{T}{3} = \frac{2\pi}{3}\frac{T}{T_n} \tag{15}$$

Thus the load function is described by F and T, and the system is described by T_n and the static deflection. Some maximum deflections will now be studied for different ratios T/T_n of load to system. For example, Figure 6b shows the phase plane and the displacement history for the case $T/T_n = 0.5$. This means that the pulse duration is half the natural period of the system; or, conversely, that the natural frequency of the system is low with respect to the load duration. This is called a soft system. Note, point A in Figure 6b, that the maximum displacement is 1.5 times the equivalent static displacement under load F. In the same way, the cases for $T/T_n = 1.5$ and $T/T_n = 3.0$ are shown in Figures 6c and 6d, to determine their respective maxima. Note these are point B, $X/X_s = 1.0$, and point C, $X/X_s = 2.0$. In this way, the phase planes for many values of T/T_n are studied, and the maxima are plotted in the response spectrum of Figure 7. For a given load function, response spectra of this type allow one to judge how to design the foundation system. For example, note in Figure 7 that response ratios less than unity are possible only at very low ratios of T/T_n, for example, $T/T_n < 0.2$. This means that the natural period should be 5 or more times the load duration, for this load shape; or that the natural frequency (cps) should be:

$$f_n < \frac{200}{T}$$

where T is in milliseconds. Thus, if the positive phase duration were 100 ms, then the natural frequency should be less than 2 cps, for example. Natural frequencies of 2 cps are not common in foundations, so one might alternatively try to design for a response ratio of unity, at, for example, $T/T_n = 1.5$, point B. This means that the natural frequency must be,

$$f_n = \frac{1500}{T}$$

Thus, for T = 100 ms, f_n = 15 cps, which is a more reasonable frequency for a foundation.

The rotational modes have very small damping, as shown in Figure 3, so that the spectrum techniques just described are helpful for design. The translational modes, however, usually have appreciable damping; and the beneficial effects of that damping must be considered. For example, the response spectrum for 50 percent damping is shown dotted in Figure 7. Note that the maximum responses are all less than the corresponding undamped case. Furthermore, response ratios less than unity can be obtained for,

$$f_n = \frac{700}{T}$$

Thus for T = 100 ms, a response frequency less than 7 cps would be required. Massive, large foundations can be designed with such low frequencies.

Note in Figure 7 that for lightly damped systems, such as rotational modes, small misestimates of natural frequency can lead to large misestimates of response. For heavily damped systems, such as translational modes, large misestimates can lead only to trivial misestimates of response.

It is interesting to compare the response spectra of Figure 7 to the spectra one would obtain by assuming the transient loads were harmonic. This harmonic, or steady-state, analysis is usually made by some arbitrary conversion of transient to harmonic. For example, the triple-peaked load function of Figure 7 might be considered as half of one cycle of a sine wave, shown schematically in Figure 8. Under this arbitrary assumption, the steady-state response spectra can be drawn from

standard published curves[10]. These are shown dotted in Figure 8 for the undamped and the 50 percent damping cases. It is important to note that the steady-state solution can be either conservative or unsafe for the undamped case; but that for the heavily damped case, the steady-state solution might be adequate. These conclusions apply only for this particular triple-step load function. Experiences with other load functions show that, in general, the steady-state approximation is inadequate for design purposes; and that the response spectra must be constructed even to determine if the steady-state is conservative or unsafe. Considering the simplicity and ease of transient analysis by the phase-plane technique, there is little to recommend continued use of so-called equivalent harmonic methods for transient loads.

The effects of soil stiffness, taken as shear modulus, are shown in Figure 9. The two curves, for no damping and 50 percent damping, were constructed by starting with the ratio $T/T_n = 0.5$ at a soil shear modulus of 10,000 psi. This is an arbitrary starting point to show effects of stiffness. Implicitly, the ratio T/T_n is different for each shear modulus, because the natural frequency of the foundation is different for each shear modulus. For example, the ratio T/T_n is 0.353 at G = 5,000 psi for this case.

Figure 9 shows that stiffer soils are indeed beneficial, especially in the lower range of stiffnesses. In addition, the figure shows the benefits of damping. For example, a system with 50 percent damping would have the same response (here, 1.0) at a shear modulus of 8,000 psi, as would an undamped system at a shear modulus of 17,000 psi. It is perhaps more meaningful to consider the effects of damping for a given soil site. For example, at an average soil stiffness of 6,000 psi, a heavily damped system will yield but slightly more than half the response of a lightly damped system.

The effects of foundation geometry are shown in Figure 10. For that figure, the ratio T/T_n was taken as 0.5 at R/H = 3. The curves for zero damping and 50 percent damping are shown as solid lines. Notice that as R/H increases (large radius R, or small height H), the dynamic displacement decreases. Thus, in general, low and wide foundations are beneficial, provided the slab still remains rigid with respect to the soil stiffness.

The two constant dampings in Figure 10 are instructive for studies of effects, but they do not represent the behavior of a foundation. Recall Equation 5, which shows that damping varies with the ratio R/H. This effect must, of course, be considered in a foundation design. The behavior of a typical foundation is shown by the dashed line in Figure 10: at low values of R/H, the damping is low, and the resulting displacements are high; but at high values of R/H, the damping is high, and the resulting displacements are quite low. Thus, for an actual foundation (dashed line), the benefits of maximizing the ratio R/H in design can be appreciable, and are much greater than for a constant-damping system.

These parametric studies have shown that response is minimized by stiff soils and by thin, broad foundations. It is necessary to reinforce such a foundation generously so that rigidity is maintained. Furthermore, these studies have shown that harmonic analysis of transient problems is not only incorrect, but is also unnecessary if available transient-analysis techniques are utilized.

V. A Design Case History

Designs of foundations for transient loads are encountered with impact-testing machines, forge hammers, car shredders, and many modern heavy manufacturing processes. It is unfortunate that a large part of the soil dynamicist's practice is spent in redesigning transient-load foundations which were either designed by rules-of-thumb or by harmonic methods.

A recent design case arose because a research laboratory proposed to build an unprecedented impact-testing machine. The machine is designed to generate a 1,600,000 pound, half-sine pulse for 8 ms. The research group was at the same time planning a new laboratory which would include the manufacturing and proof-testing facilities within one building. It was required that the manufacturing processes not be disturbed by the proof-testing, and that the building itself not be harmed by the impact tests.

The soil at the proposed site consisted of a medium dense to dense sand. In order to obtain the necessary soil properties for the transient response analysis, a seismic survey and extensive laboratory testing were performed. Included in the laboratory testing program was a series of cyclic-loading triaxial tests, performed to augment and substantiate the seismic survey, and to account for construction variables such as changed grade. Also, vibration instruments were used during the seismic survey to determine the soil attenuation parameters[8] to be utilized in the isolation studies.

Many critical design problems arose, but only one will be discussed here to show the techniques and benefits of transient-load analysis. This particular problem arose because the test pad for proof-testing the large machine was to be located directly adjacent to a building footing, as shown in the insert to Figure 11. The question arose as to whether the impact from the machine would distress the footing and adjacent wall.

Figure 11 shows the essence of the problem and the analysis. The impact load, 8,000 pounds decelerating to a maximum of 200 g's, would be applied to the top of the test pad. This is shown as sketch 1 on the figure. The test pad, having mass, stiffness, and damping, will respond to the input pulse. That response is calculated by the phase plane, shown in sketch 2. From the phase plane, the displacement (white arrow) or velocity (black arrow) histories are constructed, as has been described above. The velocity-time history can be differentiated, at least in the early portions, to obtain the acceleration-time history (black arrow) for the test pad. Because the slab is essentially a rigid body, these accelerations actually represent forces or stresses felt by the slab and transmitted into the soil. These stresses undoubtedly attenuate as they are transmitted through the soil from the test pad to the footing; however, for conservatism, no attenuation was considered in this case. Under this assumption, the output from the test pad is directly the input to the footing.

The footing and wall are a system with mass, stiffness, and damping. That system will respond to the input load, and that response is calculated by the phase plane for this system, as shown by sketch 3 in Figure 11. The velocity-time history (black arrow) of the footing is constructed from the phase-plane. The early portion (only) of that velocity-time history is then differentiated to obtain the maximum acceleration of the footing-wall system, as shown in sketch 4. The result, about one-quarter of g, can be considered in the design of the footing and wall so that substantial distress can be avoided.

Clearly, transient analysis has been carried quite far in this case history. Two phase planes and two differentiations certainly leave much room for error. The end result, however, is an estimate of considerable value because it brackets the problem: the analysis shows that the footing motions are probably less than about 1 g, but more than about 0.1 g. Thus, the computed value, 0.25 g's, indicates that special design is required, but that the design is possible. Depending on the uncertainties involved and the type of wall framing used, the structural analyst would probably want to use something greater than the computed 0.25 g's for his design.

It is interesting to note that a harmonic analysis of this problem would have yielded lower design values, perhaps on the unsafe side.

VI. Conclusions

These results show that harmonic analyses are not adequate for an optimized design of a foundation subjected to transient loads. The transient methods are, however, so easy to use that they can be applied about as economically as can the harmonic methods.

The parametric studies show that response depends strongly on soil stiffness, and perhaps more strongly on the amount of damping that can be designed into the foundation. Both stiffness and damping are enhanced by using low, broad foundations. Care must be taken, however, to reinforce the foundation so that it remains rigid with respect to the soil stiffness.

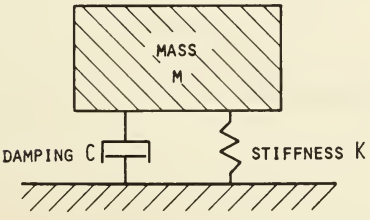


FIG. 1 - MODEL FOR FOUNDATION MOTIONS

References

1. Richart, F. E., Jr.: "Foundation Vibrations," Trans., ASCE, Vol. 127, Part I, 1962.

2. Barkan, D. D.: Dynamics of Bases and Foundations, McGraw-Hill, 1962.

3. Lysmer, J., and F. E. Richart, Jr.: "Dynamic Response of Footings to Vertical Loading," Proc. of J. of Soil Mechanics and Foundation Division, ASCE, Vol. 92, SM-1, January 1966.

4. Whitman, R. V.: "Analysis of Foundation Vibrations," Vibration in Civil Engineering, Butterworths, London, 1966.

5. Lysmer, J.: Vertical Motion of Rigid Footings, Ph.D. thesis, Univ. of Michigan, 1965.

6. Hsieh, T. K.: "Foundation Vibrations," Proc., I.C.E., paper 6571, London, 1962.

7. U.S. Army Engineers Waterways Experiment Station: Field Vibratory Test Data, Tech. Report 3-632, Rep. 1, Vicksburg, Mississippi, 1963.

8. McNeill, R. L., B. E. Margason, and F. M. Babcock "The Role of Soil Dynamics in the Design of Stable Test Pads," Amer. Inst. Aeronautics and Astronautics, Vol. Tech. Papers, Guidance and Control Conference (Minneapolis), New York, 1966.

9. McNeill, R. L., and B. E. Margason: Modern Foundation Design for Dynamic Loadings, paper as presented at the SEATO Conference on Soil Engineering, Bangkok, April 1967.

10. Jacobsen, L. S., and R. S. Ayre: Engineering Vibrations, McGraw-Hill, 1958.

11. Margason, B. E., J. A. Barneich, and F. M. Babcock: "Optimized Design of Ground-Founded Microprecision Slabs," Amer. Inst. Aeronautics and Astronautics, Vol. Tech. Papers, Guidance and Control Conference (Huntsville), New York, 1967.

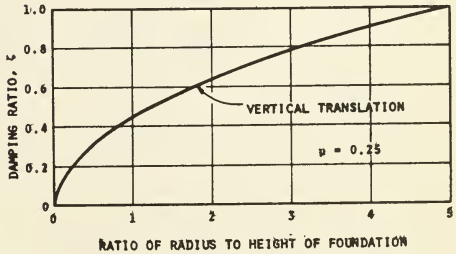


FIG. 2 - EFFECT OF FOUNDATION GEOMETRY ON DAMPING
(CIRCULAR, SOLID CONCRETE FOUNDATION)

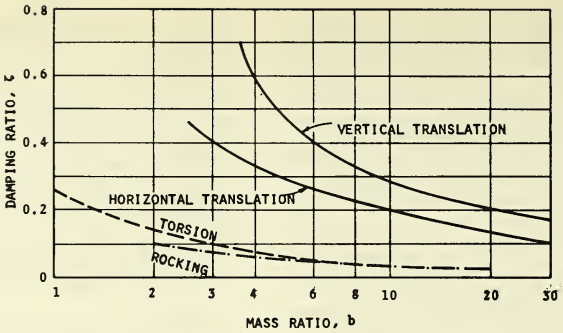


FIG. 3 - DAMPING RATIOS
(REFERENCE 4)

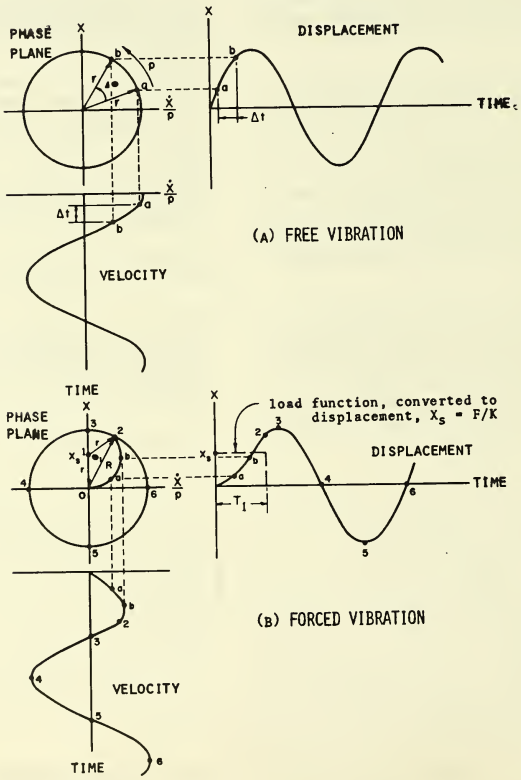


FIG. 4 - PHASE PLANE, UNDAMPED SYSTEM

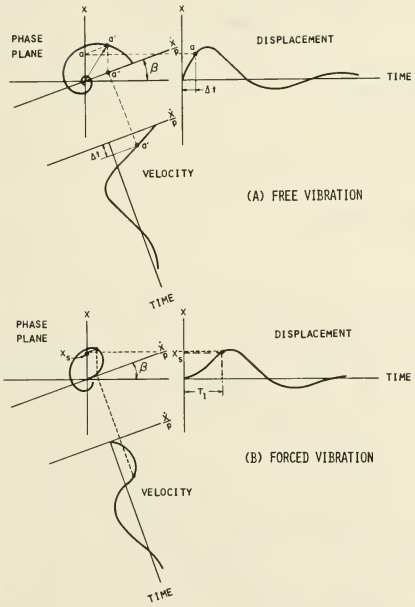


FIG. 5 - PHASE PLANES, DAMPED SYSTEM

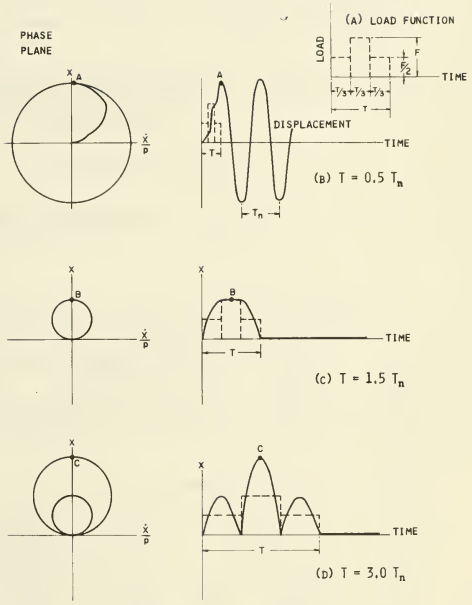


FIG. 6 - MAXIMUM RESPONSES

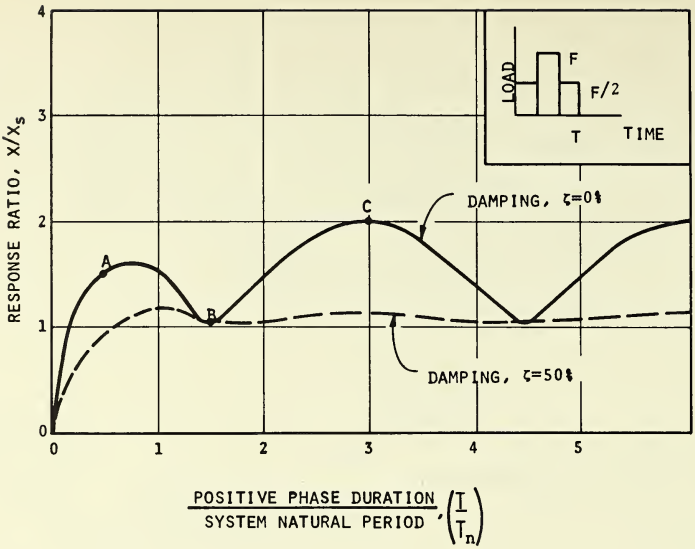


FIG. 7 - TRANSIENT RESPONSE SPECTRA

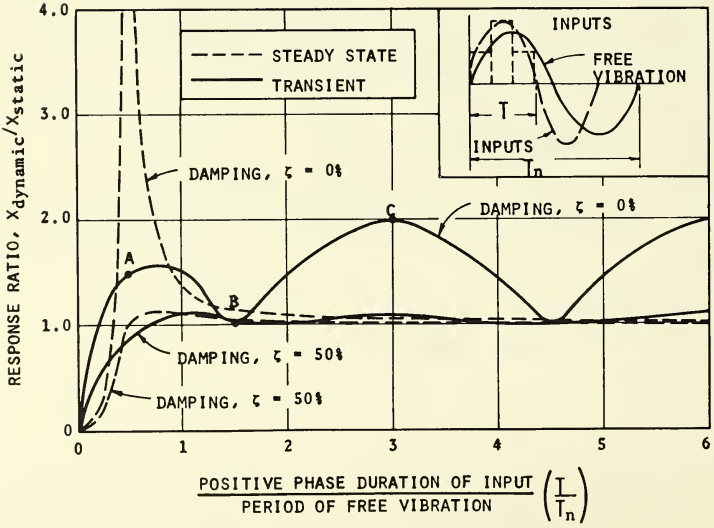


FIG. 8 - MIXED RESPONSE SPECTRA

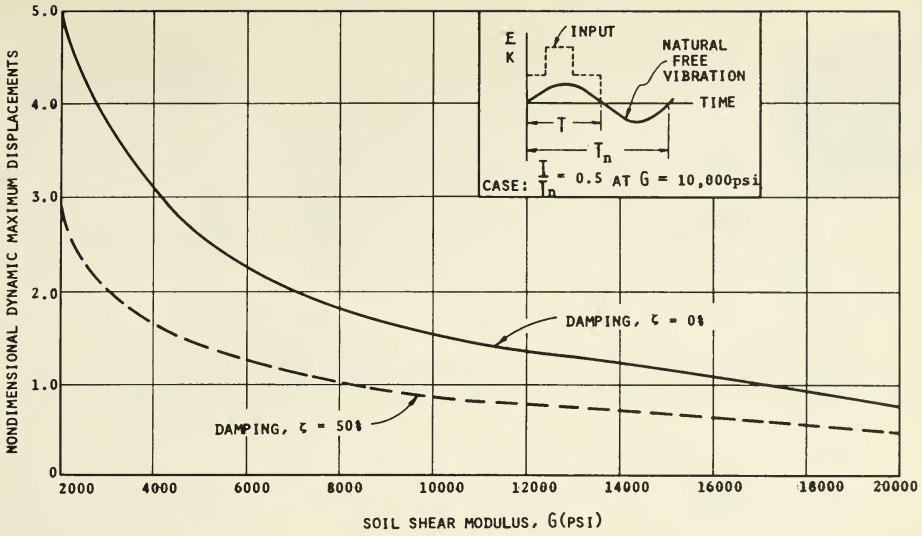


FIG. 9 - EFFECT OF SOIL STIFFNESS AND DAMPING

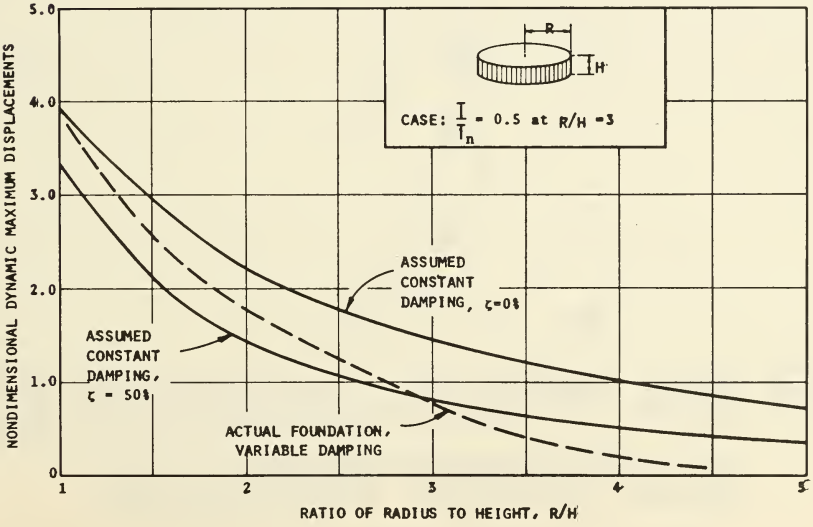


FIG. 10 - EFFECTS OF FOUNDATION GEOMETRY

67-550

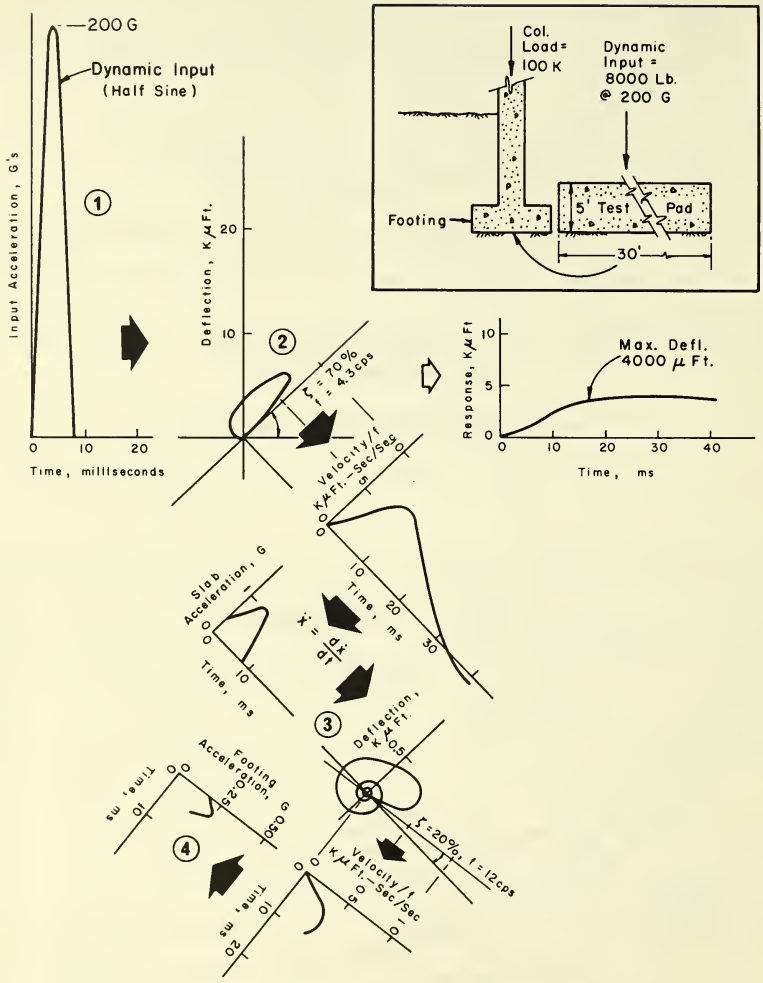


FIG. 11 - TRANSIENT RESPONSE ANALYSIS

No. 67-551

A STRAPDOWN STAR TRACKER FOR SPACE VEHICLE ATTITUDE CONTROL

by

MORRIS M. BIRNBAUM
Hycon Mfg. Company
Monrovia, California
and
PHIL M. SALOMON
Jet Propulsion Laboratory
Pasadena, California

AIAA Paper
No. 67-551

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.03, 8.04, 8.08

A STRAPDOWN STAR TRACKER FOR SPACE VEHICLE ATTITUDE CONTROL

Morris M. Birnbaum, Hycon, Monrovia, California
Phil M. Salomon, Jet Propulsion Laboratory, Pasadena, California

ABSTRACT

The strapdown star tracker is capable of locating and tracking a wide range of preselected target stars within an 8-degree field of view. The star tracker has a 10 Hertz tracking bandwidth and achieves a pointing accuracy of 12 arc-seconds for a +3 A0 star by the use of unique methods of scanning and error signal determination. An image dissector is used as the sensor, and allows electronic scanning of the field of view without any moving parts. Additionally, the small instantaneous field of view of the image dissector allows star acquisition and tracking in a relatively high background of ambient light. Different size scan patterns are used for acquisition and for tracking. In each case, the scan pattern is a rosette whose geometry and scan rate is optimized to provide the highest signal-to-noise ratio consistent with the system bandwidth requirements. Descriptions of the star tracker optical and electronic systems and an analytical determination of systems performance are presented. Comparisons of theoretical and measured performance characteristics are also presented, and are in close agreement. The star tracker is intended to be used as a primary attitude determining system for research rockets and space vehicles. High reliability, low weight, and low power consumption make it compatible with the requirements of most vehicles.

INTRODUCTION

During the past decade, a large number of devices have been built for establishing the attitude of a vehicle in space. Star trackers, as a special class of attitude sensing devices, have undergone intensive development to improve their accuracy and reliability. As a result of this effort, star trackers are finding widespread application in the guidance of research rockets, space vehicles, high-altitude research balloons, and military weapons systems.

The need for a strapdown star tracker for Aerobee rocket guidance led to the development of the Hycon HST-421 Star Tracker. The primary requirements imposed by the parent vehicle were that the star tracker have the capability of locating a preselected target star within a wide optical field of view and of tracking the star to a very high accuracy. Additionally, vehicle dynamics required a rather fast system response, both in locating and tracking the target star. Reliability considerations ruled out the possibility of employing any moving mechanisms. Thus, the design of the star tracker had to satisfy several conflicting requirements.

A large star tracker field of view is desirable to relax the pointing accuracy requirements of the vehicle coarse acquisition system. However, for high pointing accuracy, a small field of view is needed. A small field of view, once it contains the target star, allows the use of a simple acquisition and track system since the problems associated with sky background and sensing wrong stars are minimized. Large collecting optics are desired because they enhance the star signal, but they are undesirable from the standpoint of weight and volume.

The resultant star tracker system design fulfills the performance requirements by employing a dual-mode rosette scan pattern, a unique error detection scheme, and a highly efficient optical system. The star tracker is capable of locating and tracking a +3.0 magnitude A0-class star to an accuracy of 12 arc-seconds within an 8-degree field of view. This tracking precision is achieved while maintaining a system bandwidth of 10 Hertz.

SYSTEM DESCRIPTION

The Hycon star tracker uses an image dissector tube as the sensor. The field of view is scanned and the target star is located. The star can be tracked off axis, or the error signals generated can be used to orient the vehicle or platform so that the target star is centered on the star tracker optical axis. In either case, the star tracker and its associated electronics use less than 10 watts of power, require +28 vdc ±10 percent, and weigh less than 10 pounds. The star tracker and its electronics can be packaged in a cylinder 5 inches in diameter and less than 11 inches long.

The star tracker has no moving parts. It senses the angular error between its optical axis and a line to the target star by electronically scanning its field of view and generating X- and Y-error signals proportional to the angular error. The use of an image dissector to scan the field of view, as compared with a four-quadrant type tracker, allows acquiring and tracking a star in a relatively high background of ambient light.

The design features unique to the Hycon star tracker are its dual mode scan patterns and its method of deriving error signals.

An 8-degree field of view is scanned with a sinusoidal rosette pattern as shown in figure 1. When the star to be

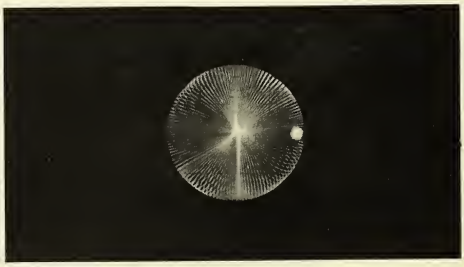


Figure 1. Acquisition Scan with Star at Edge of Field of View.

tracked is intercepted, the star tracker generates cartesian error signals by sampling and holding the X- and Y-voltage values of the sweep deflection waveform at the instant of star intercept. The magnitudes and polarities of the error signals define the position of the target star within the scanning raster. The error signals are used to drive the star towards the optical axis. When the star is near the star tracker optical axis, the acquisition rosette scan pattern is changed to a track scan pattern of much reduced diameter and reduced number of petals. The track scan, figure 2, covers just the photocathode area around the star. Due to its small size, dwell time of the track scan pattern on the star increases significantly, the field of view being scanned is small, hence, the signal-to-noise ratio increases, and tracking accuracy becomes very high.

Figure 2. Track Scan with Star at Center of Field of View.

The method by which the star tracker error signals are generated involves sampling and holding the voltage value of the sweep deflection current waveform at the instant of star intercept. Since the voltages obtained in this fashion are directly related to the instantaneous position of the target star, a cartesian set of error signals is generated which defines the position of the target star on the image dissector photocathode. The primary appeal of this means of error detection is the fact that the ability to derive error signals becomes independent of the form of the scan pattern utilized, or of any distortion in that selected scan pattern. Regardless of the characteristics of the scan pattern, the error signals always represent the resolved components of the error vector from the null axis to the target star position.

A 76-millimeter focal length, T/1, refractive optical system is used. With these optics, reliable tracking of guide stars of +3.0 magnitude or brighter is assured. System bandwidth in the tracking mode can be as high as 10 Hertz. Pointing accuracy for a +3.0-magnitude star is 12 arc-seconds, rms; with lower system bandwidths and/or brighter stars, pointing accuracy is increased.

The major system characteristics of the Hycon HST-421 Star Tracker are as follows.

a.	Field of view:	8 degrees total.
b.	Sensor:	Image dissector tube, electrostatically focused, magnetically deflected.
c.	Photocathode:	S-20, S-1, S-4, S-11, and other photocathodes available.
d.	Scan pattern:	Dual mode rosette.
e.	Dimmest star tracked:	+3.0 magnitude.
f.	Tracking accuracy:	12 arc-seconds for a +3.0 star, 7 arc-seconds for a +1.0 star.
g.	System bandwidth:	Up to 10 Hertz.
h.	Input voltage requirements:	+28 vdc ±10 percent.
i.	Power consumption:	10 watts.
j.	Weight:	10 pounds.
k.	Volume:	5 inches in diameter by 11 inches in length.

SENSOR

The advantages of an image dissector over other types of sensors are its simplicity and reliability. The tube does not have an electron gun or a thermionically heated cathode. Its aperture plate is a mechanical component, and can be mechanically located and positioned like any optical element. Because the image dissector does not employ an electron gun to scan its light sensing surface, optimized nontelevision type scan patterns can be used for acquisition and for tracking, without fear of damaging the tube photocathode surface. The operating principle of the image dissector, whereby only a small portion of the photocathode is being scanned at any instant, reduces dark current to an insignificant level, and allows sensing stars in a relatively high ambient light background. The image dissector aperture passes all of the star signal, but only a small portion of the background light reaching the photocathode.

SCANNING TECHNIQUES

Both the acquisition scan and the track scan are sinusoidal rosette patterns. Rosette parameters are chosen to allow every element of the acquisition field of view, as imaged on the photocathode, to be scanned at least once per rosette. The number of rosettes per second is determined on the basis of achieving a desired signal-to-noise ratio.

A rosette scanning pattern is selected for both acquisition and tracking because of its nondirectional scanning properties, the simplicity of its scan generating circuitry, and the ability of the scan pattern to increase the star intercept sample rate as the target star moves towards the center of the scan.

The selection of a rosette scanning pattern in the tracking mode, as opposed to other commonly used scan patterns, is based on the following considerations.

Each scan petal emanates from the center or null point of the pattern; hence, the scan pattern is nondirectional with relation to the shape of the star image on which it centers. The rosette scan pattern nulling point on a star image is affected less by image aberrations such as flare, noncircular star image, etc. Also, the use of a rosette scan pattern to scan a selected field of view around a star allows a smaller image dissector aperture to be used than if an X-Y scan pattern is used. The smaller aperture produces a higher signal-to-noise ratio in the track mode, hence more accurate tracking. System noise is minimized because the rosette scan pattern generates X- and Y-axis servo error signals simultaneously. To generate a rosette scan pattern, the following cartesian equations must be produced:

$$X = A \cos 2\pi f t \cdot \sin n 2\pi f t$$

$$Y = A \sin 2\pi f t \cdot \sin n 2\pi f t$$

The number of rosettes per second is determined by f, A determines the diameter of the rosette, and n determines the number of petals in a rosette.

The equations are simulated easily with solid-state circuitry. A digital clock, running at nf Hertz, feeds a countdown network which counts down to f and f/90 degrees. The nf Hertz square wave and the f and f/90 degrees Hertz square waves are put through filters to extract only their fundamental frequencies. The frequencies nf and f are combined in a transformer coupled chopper and narrow bandpass filter circuits, as are nf and f/90 degrees, to produce voltage analogs of the rosette equations. Figure 3 shows the block diagram for a scan generation system producing 6.25 rosettes per second with 128 petals per rosette.

To generate a track rosette, frequencies are taken from points on the countdown network to obtain the desired track rosettes per second and the desired number of petals per track rosette. By using one of the acquisition scan components, i.e., nf the same in track as in acquisition, the number of filters needed is minimized.

ERROR DETECTION SYSTEM

Error signals are generated by sampling the deflection currents in the X- and Y-coils at the instant of star intercept. The deflection current magnitudes are converted to voltages by resistors in series with the X- and Y-coils respectively. Each deflection voltage is conducted to an analog gate and is gated through to a sample and hold circuit when a video pulse puts the gate in a sample state.

A schematic of the error detection system is shown in figure 4. A 2N4220 Field Effect Transistor (FET) is used as a transmission gate. The video pulse from the image dissector, suitably amplified, is used as a gating pulse to the FET. The sweep waveform is inputted to the drain of the FET. When a video pulse puts the FET in a sample state, the sweep waveform existing at that instant is conducted into the integrating capacitor and held.

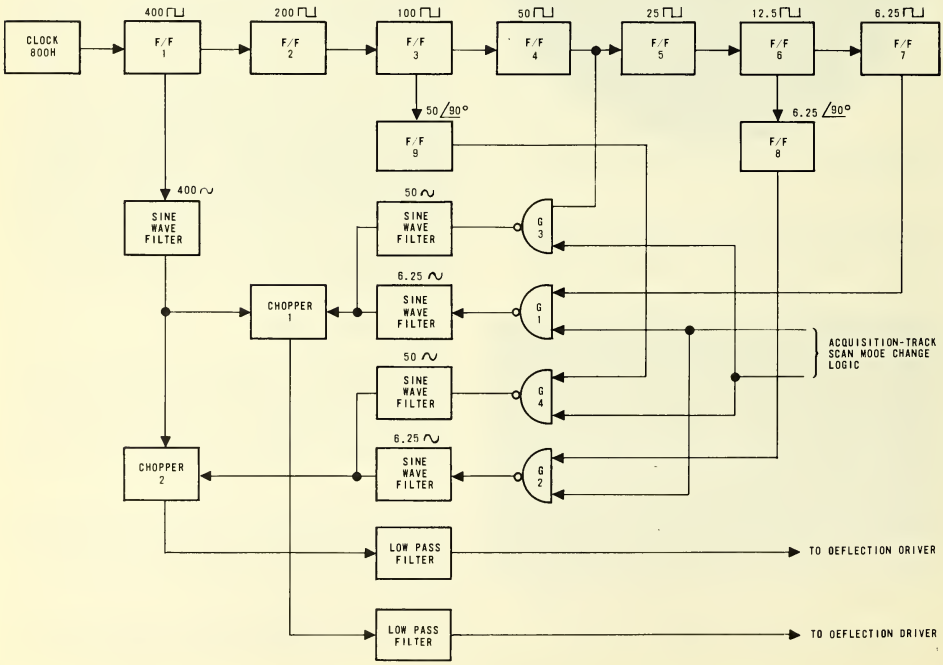


Figure 3. Scan Generator System.

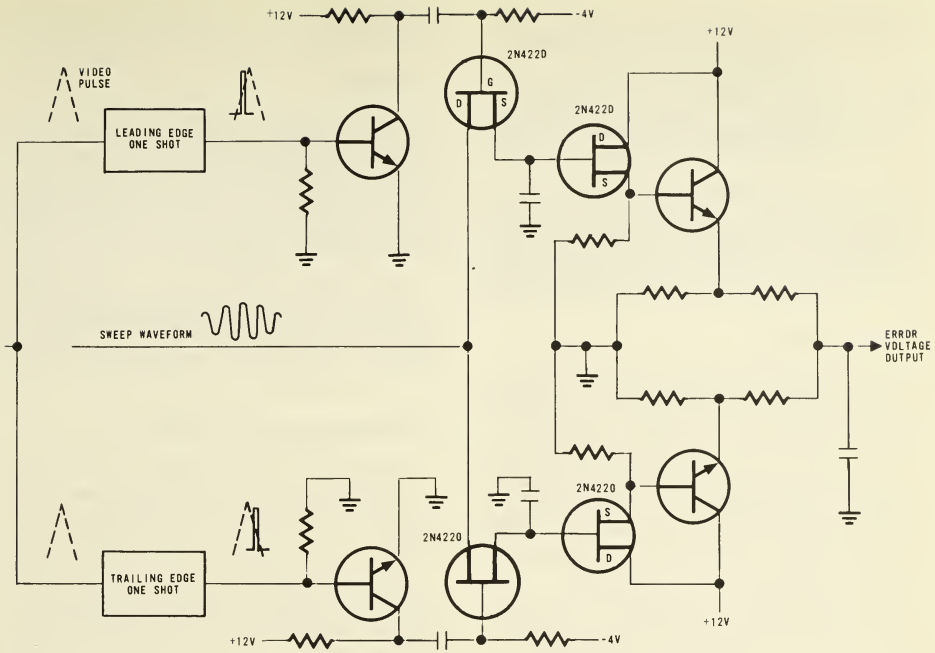


Figure 4. Error Detection, Sample and Hold Circuit.

To minimize system errors, two sample and hold circuits are used. One samples and holds the sweep waveform existing at the leading edge of the video pulse. The other circuit samples the sweep waveform existing at the trailing edge of the video pulse.

The outputs of both circuits are summed in a resistive network, and the resultant sweep waveform voltage is held in a capacitor, amplified, and used as the instantaneous position error signal existing at star intercept.

SIGNAL-TO-NOISE OPTIMIZATION

The advantage of using a small track scan, once the star is found, is because the dwell time of the scan pattern on the star is maximized. For an image dissector, the expression for signal-to-noise is given by:

$$s/n^{*} = \frac{2.25D\sqrt{P_s S_k T_a}}{\sqrt{e\left(\frac{\sigma}{\sigma-1}\right)}} \quad \text{(for nighttime conditions)}$$

where:

D = Effective lens diameter in inches.

P_s = Power density of target in watts/cm^2.

S_k = Cathode radiant sensitivity in amperes/watt.

T_a = Dissector aperture dwell time on an image element in seconds.

e = Electron charge = 1.6×10^{-19} coulombs.

$\frac{\sigma}{\sigma-1}$ = Dynode chain noise factor over and above the basic shot noise.

σ = Gain per stage, usually 3 to 5.

When the scan pattern is reduced to a track scan whose diameter is only slightly more than twice the dissector aperture diameter, T_a in the signal-to-noise expression becomes very large - about $1/3p$ of the time to generate a track rosette where p is the number of petals per rosette. With an increased T_a, s/n increases in proportion to the square root of the increase in T_a.

The pointing error is expressed by:

$$\theta_{rms} = \frac{K\alpha}{s/n\sqrt{N f}} \quad \text{arc-seconds}$$

*See appendix for derivation.

67-551

where:

K = A numerical constant depending on bandwidth and including conversion factors.

α = Dissector aperture diameter in inches.

N = Number of star intercepts by scan pattern in time period being considered.

f = Lens focal length in inches

From the equation, it is seen that θ_{rms} is inversely proportional to s/n and hence to T_a. The equation also shows that, for minimum θ, α should be small as possible, and that f should be as large as possible, implying narrow field of view for maximum tracking accuracy.

TRACKING MODES

The star tracker is capable of on-axis or off-axis tracking of selected guide stars. On-axis operation provides error signals by determining the position of the target star within the tracking raster. The tracking raster is maintained in alignment with the system optical axis. Off-axis control signals can be developed by measuring the deflection yoke current necessary to sustain off-axis tracking of the target star by the tracking raster. These currents are proportional to the displacement of the target star from the system optical axis. Track scan pattern deflection current variations are removed by low-pass filtering prior to using the deflection current measurements as a guidance control signal.

Off-axis tracking is useful when dc error signals must be generated for a slow moving servo system and the correction to be made must be accurately known. Because the off-axis system places the track scan on the star and tracks off axis, a higher s/n ratio is obtained than with on-axis tracking and the error coordinates of the star with reference to the optical axis are more accurately known.

PACKAGING

The flight configuration star tracker is shown in figure 5. Star tracker components which must be precisely aligned to each other are the lens assembly and the image dissector photocathode and aperture plate. This is done by assembling the image dissector and yoke in a sensor-optics tube and attaching the lens assembly to it. A mounting ring is screwed to the sensor-optics tube and adjusted so that its plane contains the center of gravity of the star tracker.

The electronics cylinder contains truncated pie-shaped compartments into which integrated circuit stick modules are placed. The electronics cylinder is designed so that it supports all of the weight of the integrated circuit modules with no weight being supported by the sensor-optics tube.

SYSTEM PERFORMANCE ANALYSIS

The detailed star tracker performance is presented in this section. The basic analysis is presented for the system performance with a +3.0-magnitude, A0-class target star. The analysis indicates an adequate performance margin in both detection and tracking operation.

A summary of the salient system performance parameters is presented in table 1. The analysis is based on the use of an image dissector sensor having an S-20 photocathode surface and a 0.014-inch scanning aperture. The S-20 radiant sensitivity is taken to be 0.043 ampere/watt. The optical system consists of an f/0.87 (T1) lens system. A rosette scan pattern is employed in both acquisition and tracking modes of system operation.

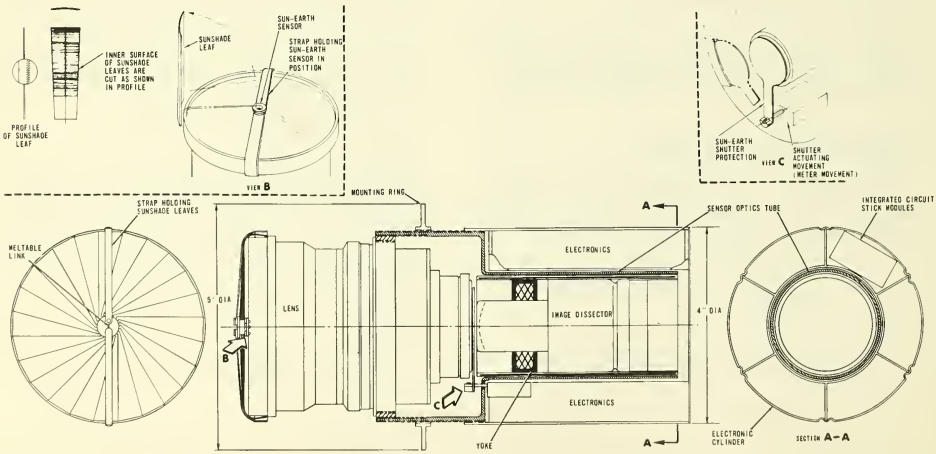


Figure 5. Flight Configuration, HST-421 Star Tracker.

TABLE 1. PERFORMANCE ANALYSIS SUMMARY

Acquisition Mode

a. Rosette scan pattern having 128 petals and 0.42-inch diameter.

b. System acquisition time equal 1 second.

c. Scan pattern rate equal 6.25 Hertz.

d. Detection s/n ratio for +3.0 star equal 5.2.

e. Optical field of view equal 8 degrees.

Track Mode

a. Rosette scan pattern having 16 petals and 0.035-inch diameter.

b. System bandwidth equal 10 Hertz.

c. Scan pattern equal 50 scans/second.

d. Detection s/n ratio for +3.0 star equal 18.

e. Effective field of view equal 40 arc-minutes.

f. Pointing uncertainty for +3.0 star equal 12 arc-seconds.

ACQUISITION MODE - PERFORMANCE ANALYSIS

Acquisition Scan Pattern Geometry. The scanned photocathode diameter of the image dissector is given by:

$$D_p = \frac{2\ fl \tan \frac{\theta}{2}}{25.4} \quad \text{(inches on the photocathode)}$$

where:

fl = focal length of lens system (76 mm).

θ = Angular field of view (8 degrees).

Using the nominal system parameter values,

$$D_p = \frac{2\ (76)(0.0699)}{25.4}$$

$$D_p = 0.42 \text{ inches}$$

The diameter of the scanning aperture for the image dissector is 0.014 inches. The number of rosette scan petals for full photocathode coverage is determined by:

$$p = \frac{\pi D_p}{\alpha}$$

For the system under consideration:

$$p = \frac{(\pi)(0.42)}{0.014}$$

$$p = 94 \text{ petals (minimum)}$$

To simplify the scan generation electronics and provide a measure of scan redundancy, a 128-petal rosette is specified.

The acquisition mode pattern rate is principally determined by the detection bandwidth requirements. The star tracker is required to generate steady-state error signals within 1 second from the instant the target star enters and remains within the field of view.

The detection time is apportioned between two serial events. The star tracker must first detect the star in the acquisition field of view, and then generate steady-state error signals within 1 second from the instant the target star enters and remains within the field of view, relating the position of the detected star to the system optical axis.

Therefore:

$$T_{acq} = T_{scan} + T_{detection}$$

For a 6.25-Hertz pattern rate in the acquisition mode,

$$T_{scan} = \frac{1}{6.25} = 0.16 \text{ second.}$$

The remaining time is apportioned to detection and generation of steady-state error signals.

$$T_{detection} = 1.00 - 0.16 = 0.84 \text{ second.}$$

The required acquisition bandwidth for steady-state (99 percent) error signals can be shown to be:

$$f_{acq} = \frac{0.733}{T_{detection}} = \frac{0.733}{0.84}$$

$$f_{acq} = 0.873$$

The 6.25-Hertz pattern rate assures a sample rate in excess of the Nyquist rate to achieve the required bandwidth of 0.873 Hertz with an adequate margin of safety.

Acquisition Detection S/N Ratio. With an acquisition mode rosette scan pattern having 128 petals and a pattern rate of 6.25 Hertz, the detection signal-to-noise ratio may be determined as follows.

The average aperture velocity for full photocathode coverage is:

$$V = \frac{pD_p}{T_{scan}} \quad \text{(inches/second on photocathode).}$$

where,

p = number of scan petals.

D_p = scanned photocathode diameter.

$$V = \frac{128\ (0.42)}{0.16} = 336 \text{ inches/second}$$

The resulting aperture dwell time, T_a, is then:

$$T_a = \frac{\alpha}{V} = \frac{0.014}{336} = 41.6 \text{ microseconds}$$

The 76-mm lens will provide an optical aperture, D, of;

$$D = \frac{fl}{T} \text{ (inches)}$$

where,

T = T number (T = 1)

$$D = \frac{76}{1} \frac{1}{25.4} = 3 \text{ inches}$$

For a +3.0, A0-class star, the resulting s/n ratio is:

$$\frac{s}{n} = \frac{2.25D\sqrt{P_s S_k T_a}}{\sqrt{e\left(\frac{\sigma}{\sigma-1}\right)}}$$

where:

$P_s = 7.1 \times 10^{-14}$ for +3.0, A0 star and S-20 photocathode.

$S_k = 0.043$ ampere/watt for S-20 photocathode surface.

$\sigma = 4$.

D = 3 inches.

$T_a = 41.6 \times 10^{-6}$ seconds.

s/n = 5.2.

The acquisition mode s/n ratio for stars of other magnitudes is shown in figure 6.

Acquisition of a target star results in error signals being generated whose magnitude and polarity determine the separation between the system optical axis and the target star. The transfer characteristics of 1 volt per degree of misalignment serve only as a means of coarse positioning (acquisition). When the star is positioned to within a predetermined distance from the system optical axis, a tracking mode is initiated, and accurate guide signals are obtained.

TRACK MODE - PERFORMANCE ANALYSIS

Track Scan Pattern Geometry. The tracking mode of system operation employs a reduced diameter rosette pattern to maximize dwell time, and thereby improves pointing accuracy.

The image dissector aperture diameter determines the minimum track scan diameter. With the selected aperture of 0.014 inch, the instantaneous field of view is

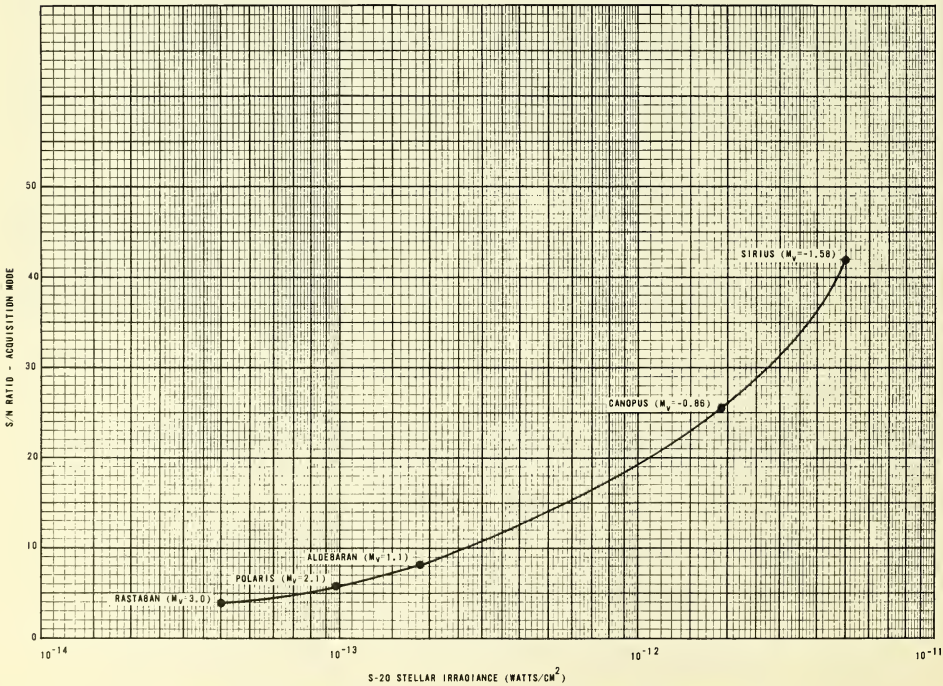


Figure 6. Acquisition Mode S/N Ratio Versus S-20 Stellar Irradiance.

16 arc-minutes. A practical minimum diameter for the track scan pattern is about 2.5-aperture diameters, or 40 arc-minutes. The resulting track scan diameter on the photocathode is then 0.035 inch. The resulting size reduction between acquisition and track scan pattern is:

$$\text{Reduction} = \frac{0.42}{0.035} = 12 \text{ times}$$

Since the track scan diameter is reduced, the number of petals required for full coverage may also be decreased. As before, the required scan petals are given by:

$$p = \frac{\pi D_t}{\alpha} = \frac{(\pi)(0.035)}{0.014}$$

$$p = 7.9 \text{ petals}$$

Again, to simplify the scan generation circuitry and also provide a measure of scan redundancy, select 16 petals.

Track Scan S/N Ratio. A pattern of 50 Hertz is selected to provide an adequate sample rate for a 10-Hertz system bandwidth. Under these conditions, the aperture velocity is:

$$V = \frac{pD_t}{T} \frac{16\ (0.035)}{0.02} = 28 \text{ inches/second}$$

The resulting aperture dwell time is:

$$T_a = \frac{\alpha}{V} = \frac{0.014}{28} = 0.5 \text{ millisecond}$$

The s/n ratio for a +3.0, A0 star is:

$$s/n = 4.88 \times 10^9 D \sqrt{P_s S_k T_a}$$

$$s/n = 4.88 \times 10^9 \times 3 \sqrt{7.1 \times 10^{-14} \times 0.043 \times 0.5 \times 10^{-3}}$$

$$s/n = 18$$

The track scan signal-to-noise ratio for other stars is shown in figure 7.

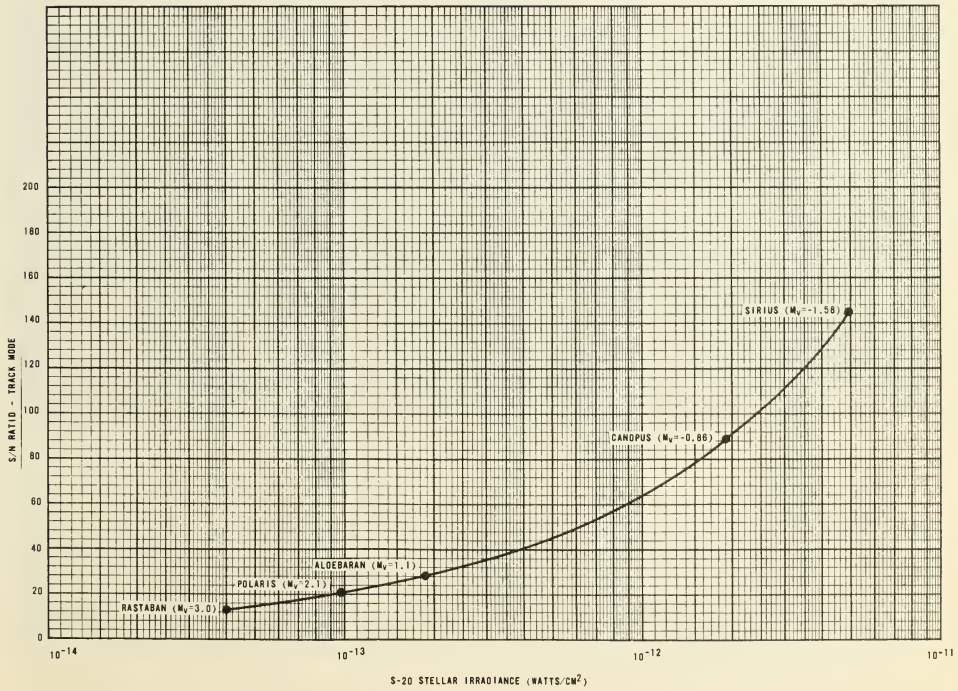


Figure 7. Track Mode S/N Ratio S-20 Stellar Irradiance.

67-551

The pointing uncertainty is then:

$$d_{rms} = \frac{p\, D_t}{s/n} \times \frac{\pi T_a}{1.6T} \times \frac{1}{\sqrt{N}}$$

$$d_{rms} = \frac{16\,(0.035)}{18} \times \frac{\pi 0.5 \times 10^{-3}}{1.6\,(0.02)} \times \frac{1}{\sqrt{80}}$$

$$d_{rms} = 0.000171 \text{ inch on photocathode.}$$

$$\theta_{rms} = 12 \text{ arc-seconds angular uncertainty for a +3.0, A0 star.}$$

The track mode pointing error for other stars is shown in figure 8.

COMPARISON WITH LABORATORY MEASUREMENTS

The actual system performance was found to agree quite closely with the predicted system characteristics. In general, measured system performance was within the measurement error as defined by instrumentation and/or simulation of target star magnitudes.

With a system transfer function of 8 millivolts/arc-second for a plus one magnitude star, the predicted angular pointing uncertainty is approximately 56 millivolts rms. The measured pointing uncertainty was found to average about 50 millivolts rms throughout the major portion of the track field. While both numbers are subject to errors due to data and measurement uncertainties, the agreement is close enough to give confidence in the analytical techniques employed. The measured system transfer curve for tracking mode is shown in figure 9. The linearity is quite good over the largest portion of the track field of view.

The system dynamic range is 10 Hertz. This was established by means of a test fixture employing a collimator and a set of mechanically rotated optical wedges. By varying the speed of wedge rotation, the apparent sinusoidal motion of the target star was caused to vary. The amplitude of the system output voltage was found to have decreased by 3 decibels when the target star image was executing sinusoidal motion at a 10-Hertz rate. An oscilloscope trace of the error output of the star tracker is shown in figure 10. A photograph of the dynamic test facility is shown in figure 11.

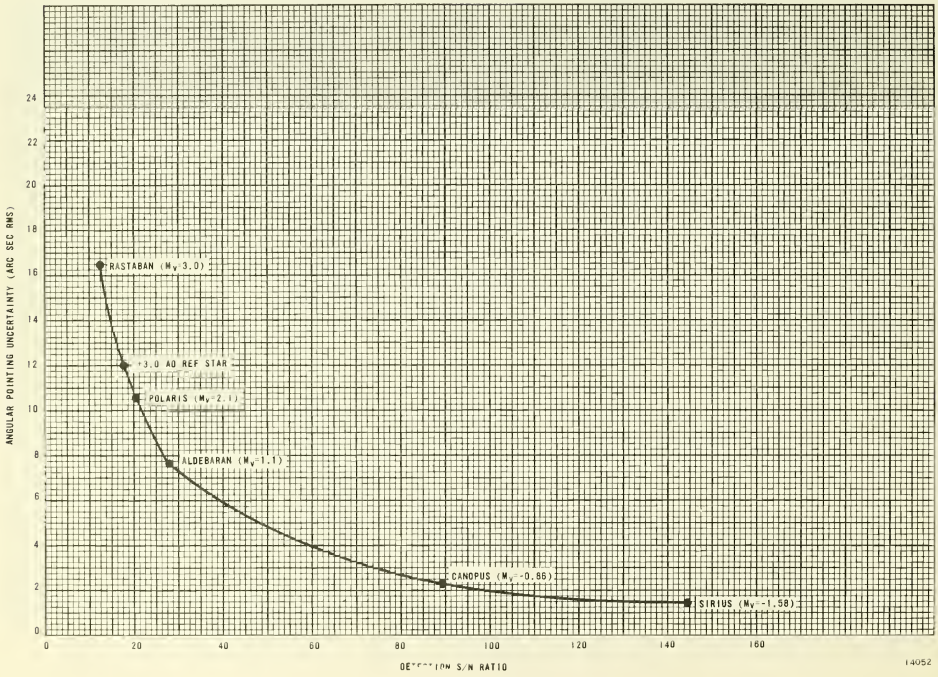


Figure 8. Track Mode Pointing Error Versus Detection S/N Ratio.

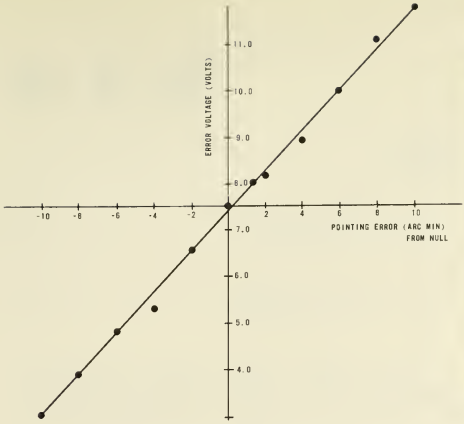


Figure 9. System Transfer Function, Track Mode, Azimuth Axis.

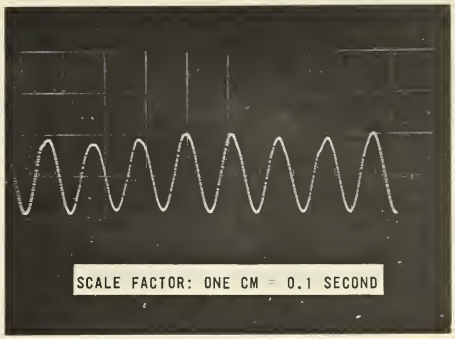


Figure 10. Dynamic Tracking Output Error Signal.

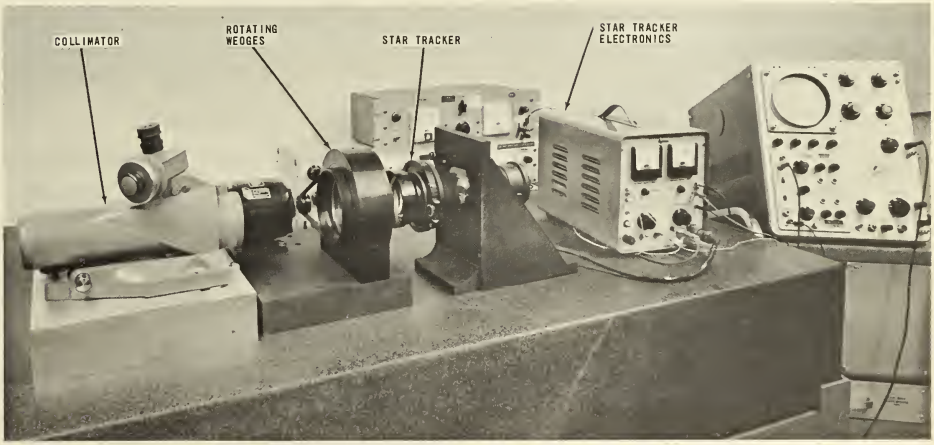


Figure 11. Test Setup to Measure Servo Bandwidth.

CONCLUSIONS

The implementation of a dual mode scan system with an image dissector allows designing a star tracker with a wide acquisition field of view and a high pointing accuracy. The use of an error sensing system which averages the position of the star pulse produces a star tracker system whose performance is close to the optimum expected as defined by signal-to-noise and root-mean-square pointing error equations.

The star tracker can be used for star field mapping with virtually no change in the basic design. All that is required is to inhibit track mode changeover. Since the acquisition scan completely covers the field of view, all stars lying within the field, above some preset threshold, will be detected. The X- and Y-coordinates of each star will be established because of the error detection techniques, i.e., using the star pulse and sampling the instantaneous value of the yoke current deflection waveform. Star position coordinates are available from the X- and Y-error outputs of the system. The format of the output is an X- and Y-error voltage which is held in each sample and hold circuit respectively until updated when another star is sensed in the field of view. The error outputs are analog voltages referenced to the star sensor

optical axis. By outputting the X- and Y- voltages to a storage network, a series of star positions can be recorded or telemetered.

APPENDIX

DERIVATION OF IMAGE DISSECTOR S/N EQUATION

Consider a target star having a power density of P_s watts/cm^2. The input power to a detector having a collecting aperture of D centimeters is:

$$P_t = \frac{P_s \pi D^2}{4} \text{ (watts)}$$

The signal current due to star detection is:

$$I_{ks} = P_t S_k \text{ (amperes)} \cdot$$

where,

S_k = cathode radiant sensitivity (ampere/watt).

The image dissector anode current is then:

$I_{as} = P_t S_k A$ for a dissector current gain, A.

The anode noise current is given by:

$$I_{an} = \sqrt{\frac{e P_t S_k A^2 (\sigma/\sigma-1)}{T_a}}$$

where,

e = electron charge in coulombs.

T_a = dissector aperture dwell time in seconds.

$(\sigma/\sigma-1)$ = corrective term for secondary emission.

The s/n ratio is then just:

$$s/n = \frac{I_{as}}{I_{an}} = \frac{P_t S_k A}{\sqrt{\frac{e P_t S_k A^2 (\sigma/\sigma-1)}{T_a}}}$$

$$s^2/n^2 = \frac{P_t^2 S_k^2 A^2 T_a}{e P_t S_k A^2 (\sigma/\sigma-1)}$$

$$s^2/n^2 = \frac{P_t S_k T_a}{e(\sigma/\sigma-1)}$$

but,

$$P_t = \frac{\pi D^2}{4} P_s$$

Therefore:

$$s^2/n^2 = \frac{\frac{\pi D^2}{4} P_s S_k T_a}{e(\sigma/\sigma-1)}$$

$$s/n = \frac{\frac{\sqrt{\pi}}{2} D \sqrt{P_s S_k T_a}}{\sqrt{e(\sigma/\sigma-1)}}$$

$$s/n = \frac{0.886 D \sqrt{P_s S_k T_a}}{\sqrt{e \frac{\sigma}{\sigma-1}}} \text{ for D in centimeters.}$$

$$s/n = \frac{2.25 D \sqrt{P_s S_k T_a}}{\sqrt{e \frac{\sigma}{\sigma-1}}} \text{ for D in inches.}$$

BIBLIOGRAPHY

William D. Atwill, "Star Tracker Uses Electronic Scanning," Electronics, Sept. 1960, pgs. 88-91.

Patent No. 3,240,942, "Apparatus For Establishing And Maintaining A Predetermined Angular Relation To A Source Of Radiant Energy," by M. M. Birnbaum and P. M. Salomon.

Floyd V. McCanless, "A System Approach To Star Trackers," IEEE Transactions On Aerospace And Navigational Electronics, Sept. 1963, Vol. ANE-10, No. 3, pgs. 182-193.

Norman P. Laverty, "The Comparative Performance Of Electron Tube Photodetectors in Terrestrial and Space Navigation Systems," IEE Transactions On Aerospace And Navigational Electronics, Sept. 1963, Vol. ANE-10, No. 3 pgs. 194-205.

James L. Lawson and George E. Uhlenbeck, "Threshold Signals," McGraw-Hill, 1950.

Glen Quasius and Floyd McCanless, "Star Trackers and Systems Design," Macmillan and Co., Ltd. 1966.

No. 67-552

INTEGRATING THE OPERATIONAL SPACECRAFT DISPLAYS INTO THE FUTURE GUIDANCE AND CONTROL SYSTEMS

by

J. R. ACKER and R. G. RAYMOND
Honeywell, Inc.
Minneapolis, Minnesota

AIAA Paper
No. 67-552

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

8.09, 12.10

-- NOTES --

INTEGRATING THE OPERATIONAL SPACECRAFT DISPLAYS INTO THE FUTURE GUIDANCE AND CONTROL SYSTEMS

J. R. Acker and R. G. Raymond
Principal Development Engineers

Aerospace Division
Honeywell Inc.
Minneapolis, Minnesota

ABSTRACT

Manned-spacecraft have been increasing in sophistication with the result that the number of displays related to guidance and control exceed the available prime panel viewing area and tax man's ability to monitor them. A time-shared "multiformat" display medium may help to relieve this dilemma. To this end, thin-film, X-Y addressable, electroluminescent (EL) displays were evaluated and an operational, computer-driven display system built. This breadboard display incorporates a 7 by 7 inch thin-film EL screen having a resolution of 37 lines per inch. Photographs of actual displayed patterns show brightness of 5 to 10 foot-lamberts and excellent cross suppression. From a hybrid simulation it is concluded that complex patterns such as an attitude-ball format cannot suitably be displayed on a moderate resolution screen because curved lines and small symbols are excessively distorted. A flight-model display system has been devised using a 4 by 4 inch EL screen of 160 by 160 lines with size, weight and power estimates included to show what improvements might be expected in the near future.

INTRODUCTION

Guidance and control systems are becoming more complex as their respective vehicles have become larger, more expensive and more sophisticated. To date, this added complexity has resulted in the addition of more equipment on a largely functional basis. In the past this has been an effective way of keeping interfaces simple and making test and operating procedures clear cut, but a point has been reached where there is no longer prime panel space available for independent subsystem displays. For example, more than 150 different display devices are used in the Apollo Command Module. Thus, it is necessary to do two things: first, provide a multiformat display to present various subsystem functions in sequence, and second, integrate the subsystems which interface with the displays on a physical rather than just a procedural basis.

Of the display mediums available today the cathode ray tube (CRT) would be by far the most likely candidate if it were not for the problems of implosion, sensitivity, size, etc., and the fact that while it is and has been available, it has not gained wide acceptance as an airborne display device. It is apparent that to gain the ultimate acceptance a display device must be developed which uses a solid-state display medium. A very likely candidate to fulfill the needs for a multiformat solid-state display device is the electroluminescent (EL) crossed grid which is solid-state, flat, potentially low cost, implosion free, and at a reasonable stage of development.

SELECTION OF A DISPLAY MEDIUM

As the limiting factor in developing a multiformat solid-state display is the state of the art of the various display mediums, the first step is to procure or develop one. In order to accomplish this first step, a set of mandatory and desirable characteristics were established as listed below:

1. The display medium must be an X-Y addressable digital type, capable of being batch fabricated, with a good potential of high resolution, low power, high reliability and small size, along with acceptable brightness and contrast.

2. It is desirable that the display medium be potentially low in cost, provide a feasible display by 1970, be compatible with integrated circuit voltage levels, and be available today with the "mandatory characteristics" in prototype form.

These characteristics were used to select thin-film EL material over the two other display mediums considered - sintered EL material and light-emitting diodes, (injection electroluminescence). Sintered EL material was not used because the resolution (lines per inch) is low compared with thin film. The light-emitting diodes were not used because the available arrays were not large enough and the time period in which these items could provide a good display is beyond 1970.

An important consideration in selecting thin-film EL was its availability in large-sized, high resolution panels (7 inch x 7 inch - 37 line/in.) since it was deemed important to develop techniques that were applicable to X-Y addressable displays; also, the technology developed for an available EL type display would be transferable to any X-Y-type display medium.

CHARACTERISTICS OF THIN-FILM CROSSED GRIDS

Figure 1 shows a section of a crossed grid panel and details how an individual element on a

67-552

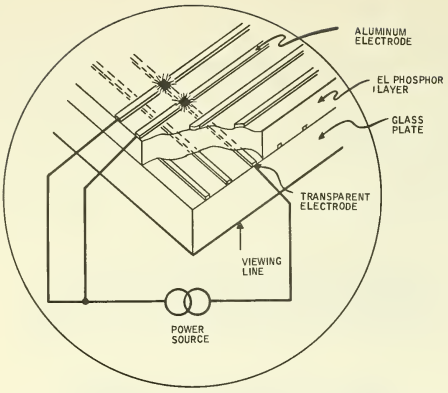


Figure 1. EL Construction View

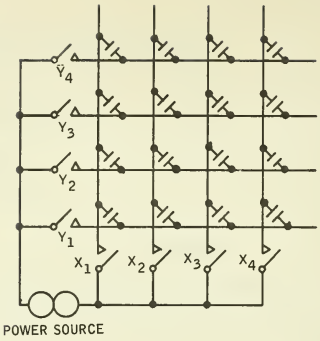


Figure 2. EL Switching Arrangement

panel is excited. The panel is built by depositing a transparent conductive layer that is photo etched to form electrodes at the required resolution and size on the glass plate. The EL phosphor is next deposited and, finally, a metallic (aluminum) electrode pattern is deposited perpendicular to the transparent electrodes. The intersecting areas of the two electrodes (with the EL sandwiched between) form discrete light-emitting elements when appropriately excited. The figure shows that the power source connections light two elements along the third transparent electrode at the intersection of the third and fourth aluminum electrodes. This type of construction results in an X-Y addressable matrix type screen that can display relatively complex patterns when information is sequentially supplied to both the X and Y axes.

There are several ways to sequence information into the display screen. Since each discrete element on the screen has a unique "X" address and a unique "Y" address the information may be supplied on an element-by-element or on a line-by-line basis, or on a partial line-by-line basis, with the constraint that the selection be limited to one "Y" address for any desired combinations of "X" addresses. Using Figure 2 as a matrix representation of the EL screen and assuming it is desirable to excite the EL elements in a diagonal line from the upper left hand corner to the lower right hand corner, the excitation of this diagonal requires that all X and Y switches be closed. It can be seen from the diagram that if all of the switches were closed at the same time the whole panel would light. To excite only those elements on the diagonal line requires that the switches be closed and opened in a sequential manner. This scanning of the panel results in each element being excited for only a small percentage of the time required to scan the entire panel. This scanning requirement causes the duty cycle of any element in the screen to be inversely proportional to the total number of lines (or elements) being excited. For this reason, and since brightness is proportional to duty cycle, it is desirable to use a line-by-line scan technique rather than an element-by-element scan technique. Figure 3 shows the voltage applied to a single element with respect to the power source waveform for a typical line-by-line scan. It can be seen that the excited element is actually being driven by voltage bursts that repeat at some relatively low frequency (30-60 Hz).

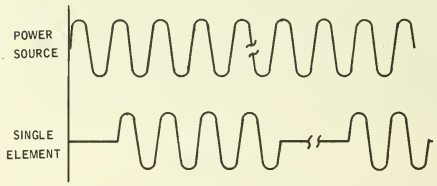


Figure 3. Drive-Source, Duty Cycle

An important feature of the thin-film EL phosphor is that the light output (brightness) of the thin film is non-linear with respect to the input voltage. This characteristic becomes quite useful in minimizing the cross talk (lighting of unselected elements) in a cross-grid display. Figure 4 is a plot of brightness versus voltage for a typical thin-film cross grid showing the rapid increase in brightness for voltages over 400 volts peak to peak.

CROSS TALK SUPPRESSION

In an EL crossed grid display when a single element is excited with a voltage V, the unselected elements in the row and column of the selected element have approximately V/2 volts applied to them. As can be seen from Figure 4, 440 volts peak to peak will produce 10 foot-lamberts of brightness, whereas, 84 percent of 440 volts will produce 0.1 foot-lamberts. The unselected elements in the row and column have $\frac{V}{2}$ or 220 volts

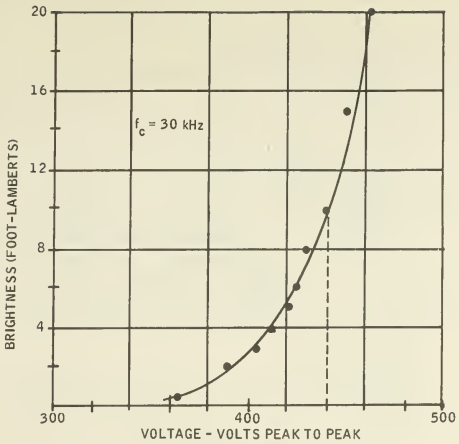


Figure 4. B-V Characteristics

across them, which, of course, will not cause light emission.

As the number of selected elements in a single row or column is increased the percentage of the applied voltage seen at the non-selected elements increases toward 100 percent of the excitation voltage. Figure 5 shows the percent of applied voltage appearing on non-selected elements as a function of the number of selected elements for several screen sizes. It can be seen from the curve for n = 250, that the unselected elements will have over 90 percent of the excitation voltage applied to them when only 5 percent of the elements in a line are excited. This will cause the unselected portion of the line to emit visible light, the brightness of which will increase as the selected elements increase, until at about 10 percent of the line selected all elements in the line will be at approximately the same brightness. The cross talk would be even greater without the excellent cross-talk suppression characteristic of thin-film EL.

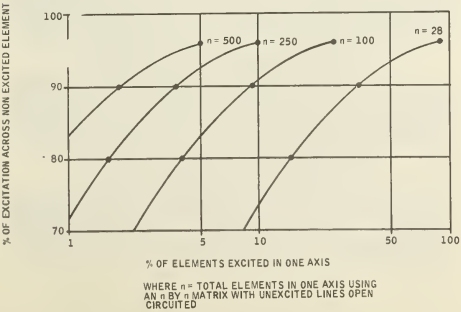


Figure 5. Cross-Talk Curves

By using the available suppression characteristics, cross talk can be completely controlled through proper design of the drive electronics. This is accomplished at the expense of power consumption by selecting termination resistors that, in effect, ground each unselected row. The value of the optimum row termination for a given level of cross talk can be determined by the following equation. The screen columns, which actually operate at a relatively high duty cycle, are terminated in a high resistance (R = 100K ohms) to minimize the power consumption.

$$R = \frac{X_c}{2\rho\,(n-1)} \sqrt{\left(\frac{\delta}{1-\delta}\right)^2} - 1 \text{ ohms}$$

where

X_c = reactance of a single-screen element

ρ = fractional value of excited elements in a row

n = number of rows

δ = fractional value of applied voltage across the unselected elements e.g., if 75 percent is permitted for 0.1 ft-L then δ = 0.75

THIN-FILM EL SCREEN EQUIVALENT CIRCUITS

Before drive circuits can effectively be designed, an accurate equivalent circuit representing the EL screen must be available. Factors that should be considered when attempting to develop equivalent circuits are as follows:

- The nonlinearities and efficiency of conversion from an electrical excitation to light emission
- The distributed properties of the matrix elements
- The impedance of the thin-film electrodes.

The screen elements formed at the intersection of any pair of crossed electrodes behave basically like ideal capacitors. At frequencies where the phosphors are efficient the effect of shunt loss terms can be neglected. Generally, the resistance of the non-transparent rear electrodes can be neglected, whereas the resistance of the transparent front electrodes can only be neglected for small screens where the conductors are all short. If resistances are lumped at strategic points along the electrodes, the screen equivalent takes the form shown in Figure 6(a). It is desirable to externally connect all electrode ends with the same designations. In this circuit the screen elements have value C and the electrode resistors value R. It can be shown that the ends of all non-excited rows are at the same potential; the same is true of the non-excited columns. Then, as far as the equivalent is concerned, these electrodes can be bussed together.

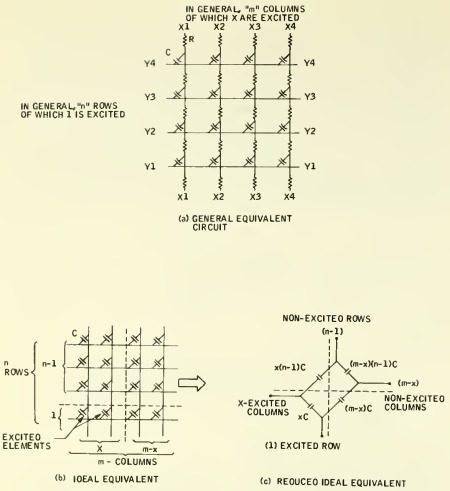


Figure 6. Equivalent Circuits

This equivalent circuit can be analyzed in order to predict cross-talk, brightness variations, screen dissipation, and total power requirements for a given set of screen parameters.

As the screen size is reduced, the effect of electrode resistance also decreases, and for small screens can be reasonably neglected. The equivalent takes the form of Figure 6 (b)(c). Typical thin-film screens have C equal to 9 pf.

DRIVE TECHNIQUES AND DEVICES

A common excitation source is used to provide the drive voltages for all electrodes along each screen axis as shown in Figure 7. Switching circuits are placed between each electrode and the common source. These circuits operate in response to logic control pulses and apply the excitation to the screen for the duration of the control pulse. The circuits associated with each line are operated in a switching mode. The switching mode circuits perform the function of simple single-pole, single-throw switches. This mode is preferred because the circuits dissipate less power, are more tolerant of load variations, and are more readily adaptable to microminiaturization.

Many switching mode devices have been proposed for use with EL screens - such as SCR's, ferroelectric devices, elemental glass bulk effect devices, photoconductors and transistors. For present applications transistor/diode circuits fill all the requirements and these circuits have been developed for use with the operational hardware described in the next subsection.

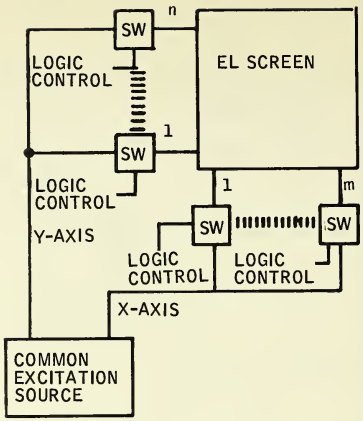


Figure 7. Excitation Scheme

The switching circuit used to control the high voltage applied to the screen elements is composed of two separate circuits. One circuit is the actual high-voltage switch and the other is the logic-level converter used to interface standard integrated circuit output levels with the high-voltage switch.

	Parts for Switch	Parts for Logic Converter
Transistors	1	2
Diodes	4	1
Resistors	0	4

OPERATIONAL MODELS

Two operational models were developed. The first, a feasibility model, utilized a small thin-film EL panel while the second model (EL7) used a large screen. Both display screens were purchased from Sigmatron, Inc., Goleta, California. Table 1 details the characteristics.

A block diagram of the small-screen display is shown in Figure 8. The display, programmed by a diode matrix, can provide duty cycle variations from 0.4 percent to 50 percent, frame rates of 15 to 63 frames per second, and pattern translation rates of 0 to 10 elements per second. Figure 9 is a photograph of the display and several patterns used to evaluate it.

Based on a successful evaluation of the small-screen display, a larger display was configured as shown in Figure 10. The digital computer (SDS 9300) provides the pattern memory, pattern computation, and sorting of the pattern information into a line form to be compatible with the EL-7 display. The basic operation of the

Table 1. EL Screen Characteristics

	Small Screen	Large Screen
Substrate Size	1.5 x 1.5 inches	10.75 x 10.75 inches
Active Area	0.85 x 0.85 inches	7 x 7 inches
Element Size	20 x 20 mils	20 x 20 mils
Gap	10 mils	7 mils
Resolution	33 lines/inch	37 lines/inch
Number of Elements	28 x 28 lines (784 elements)	258 x 258 lines (66,564 elements)
Color of Emission	5800 Å (yellow)	5800 Å

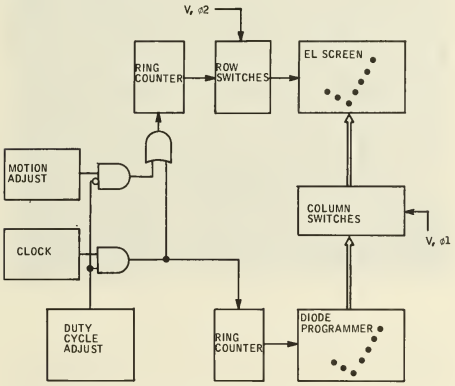


Figure 8. Small Display Block Diagram

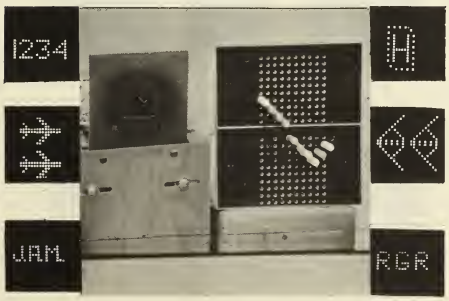


Figure 9. Photos-Patterns, Display View

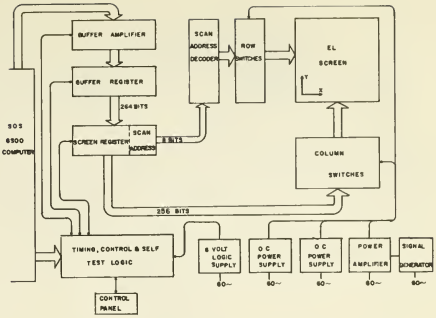


Figure 10. Large Display Block Diagram

9300/EL-7 combination is described in the following paragraph.

The pattern to be displayed is contained in memory in the computer in the form of 24-bit words placed in sequential address locations. Under control of the EL-7, 11 24-bit words are serially strobed from the computer memory into the EL-7 buffer register (Figure 10), forming a 264-bit word with the last 8 bits denoting the scan address ("Y" line) to be selected. The 264-bit word is then parallel transferred into the screen register where the first 256 bits are used to activate the column switches ("1" = ON) and the last 8 bits select a single row switch. The result at this point is a single line of the pattern displayed on the screen under control of the screen register. The buffer register is now reloaded from the computer and held in readiness for transfer to the screen register. The load and transfer process occurs at a rate that permits refreshing the entire screen 30 times per second and is synchronized to the screen excitation. The use of a buffer register and screen register[1] allows a line to be displayed while the next line is being loaded thus keeping the duty cycle on the screen for a maximum period.

PERFORMANCE

The characteristics of the EL screen when driven with modulated sine wave bursts is shown in Figures 11 through 13 for a range of voltages, duty cycles, and frame rates. Figure 11 shows the very steep voltage-brightness curve that is characteristic of the thin-film EL screen. The brightness variation between elements can be attributed to a variety of conditions including film thickness variations, variation in conductor resistivity, and poor electrical connections.

It should be pointed out that while brightness variations do exist, the visual output of the screen provides reasonable brightness uniformity and does not severely detract from the pattern acceptability. In fact, much of the visual non-uniformity can be eliminated by adjusting the voltage level and contact pressures at the screen-to-connector interface.

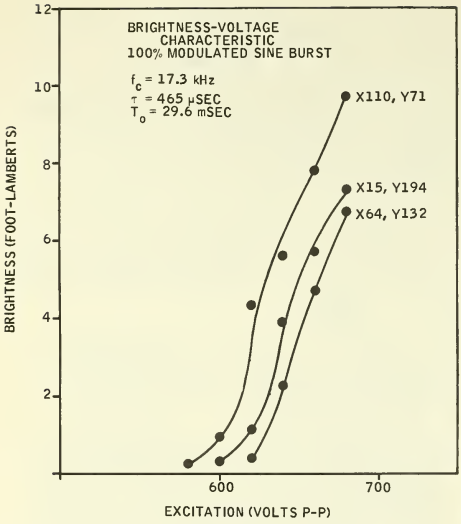


Figure 11. Screen Characteristics

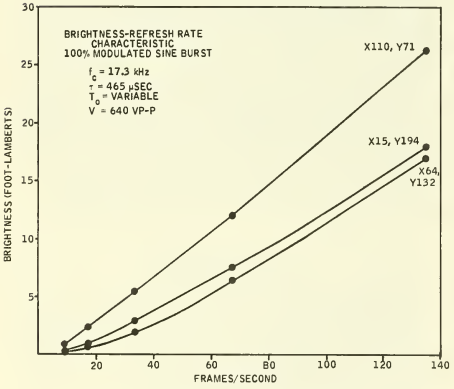


Figure 12. Screen Characteristics

Figure 14 shows the EL-7 hardware and Figure 15 contains photographs of actual formats displayed on EL-7.

Problems encountered that affected performance were the difficulty of making reliable and economical electrical connections to the closely-spaced screen conductors, excessive sensitivity of the screen to voltage transients and the dependence of brightness response time on previous brightness history.[2]

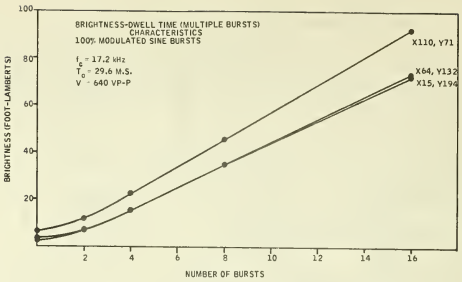


Figure 13. Screen Characteristics

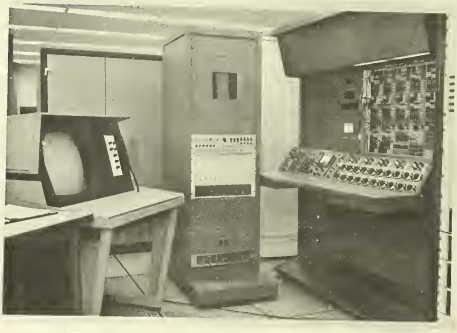


Figure 14. Hardware Photo

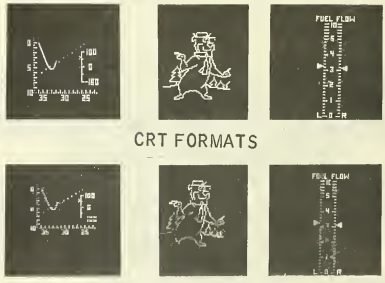


Figure 15. Formats

SYSTEM CONSIDERATIONS

Part of the development of a multiformat solid-state display necessitated the evaluation of the acceptability of a flat-plate display when applied to patterns normally presented using electromechanical devices. To accomplish this evaluation, a computer simulation program was developed using the Apollo flight director attitude indicator (Figure 16) as a display model with the ball pattern displayed on the cathode ray tube (CRT) output of the computer. In addition to investigating the acceptability of a flat-plate display, the simulation was to determine the human factors requirement[3] for a flat-plate display, study the advantages of the multi-format approach, and develop the computer software[4] necessary to apply a matrix-addressed display medium prior to the development of the EL-7 display.

Figure 16. Ball Photo

Basically, the CRT simulation operates in the following manner. Body rate commands from the hand controller (Figure 17) are processed in the computer on an interrupt basis and in turn are converted to Euler rates and integrated. The resulting integrals (Euler angles) completely define orientation of the simulated ball relative to the scope face. Size of the bezel or viewing window can be varied by operator command, thereby providing capability to "see" one half of the ball or less. Given ball attitude, the format segments in memory are tested for visibility, i. e., visible portions must be determined and projected onto the scope. These transformed coordinates (line segment endpoints) are scaled to scope coordinates and stored in one of two areas of computer memory set aside for scope commands. Line segments are filled in or drawn by employing the vector-drawing capability of the DD40 CRT system. Two areas of memory are used for storage of scope commands because simultaneous computing and output are required. Using this technique, the computation cycle time for a new attitude was 250 milliseconds and the scope was refreshed every 20 milliseconds. The results of the CRT simulation showed that a complex ball-type pattern is not suitable for display on a moderate resolution crossed grid because small markings and curved lines become distorted at a

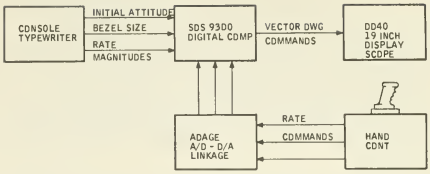


Figure 17. Simulation Block Diagram

skew angle and computation time on an element basis approaches one second on a high-speed computer.

A second mode of the simulation is one whereby the pattern is generated using discrete dots rather than line segments. The number of dots per inch can be varied to simulate the resolution of an EL cross-grid type display. This allows the creation and display of any desired pattern in a static mode with the additional feature that such items as pointers, trace histories, and characters can be programmed to provide a dynamic display. Figure 15 shows several patterns generated using this simulation. The upper photographs are of the pattern generated using the dot mode of the CRT and the lower photographs are of the pattern displayed on the EL-7 display.

FLIGHT MODEL ESTIMATES

As a basis to determine the size, weight, and power for a flight model display, a display system was devised using a 4 by 4 inch EL screen of 160 by 160 lines for a resolution of 40 lines per inch, as shown in Figure 18. A limited analysis of software techniques[5] proved useful in formulating this block diagram. Preliminary estimates indicate that the pattern and computer memories will each have a capacity of 2048

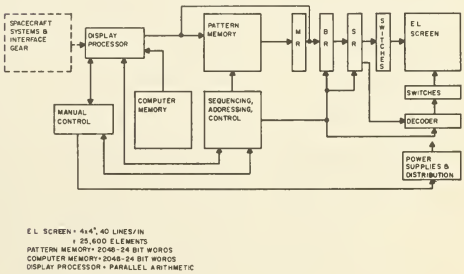


Figure 18. Flight Model Block Diagram

67-552

24-bit words. Pattern size has been limited to 50 lines with lit elements at any one time. The pattern memory will include 50 alphanumeric-type characters - for example, the alphabet, numerals 0-9 and a few selected symbols and Greek letters. Similarly, straight and curved lines (both solid and dashed), hash marks, arrows, etc., are included. These are all stored on a bit per screen element representation of the character on a 6 by 7 element grid. A separate line address table speeds identification of lit lines.

All data inputs are in digital form with some data transformations required. The processor is assumed to be similar to an ALERT computer with the following characteristics: parallel arithmetic, 24-bit word length, 89 instruction repertoire, 4 μsec add and 14 μsec multiply. Table 2 summarizes the results of estimates[6] for two time periods to show what improvements can be expected in the near future. To provide the improvements shown by the 1970-1971 time period will require the present rate of evolution of integrated circuits and memories for the display computation and control functions; this evolution can be assumed with a high degree of confidence. To attain the improvements shown for the EL switching (whose power is a function voltage) and power supplies will require more than a refinement of today's state of the art. The single most important improvement that must be made is a reduction in the operating voltage of the EL. This one item alone should ensure meeting the 1970-1971 estimates.

RESULTS AND CONCLUSIONS

The significant accomplishments of the program to date are:

- A large crossed grid display panel has been used to display a variety of complex patterns.
- The application parameters for a thin-film panel have been defined and effective solid-state switching circuits have been developed.
- The data base necessary to determine the feasibility of a flight model display has been established.
- An understanding of the computation and interface requirements of a matrix addressed display has been developed.
- A versatile tool (EL-7) has been developed which provides a means to evaluate EL crossed grids.
- A technique to model various display mediums using a computer-driven CRT was developed.
- Problems and constraints associated with thin-film EL crossed grids have been identified.

Table 2. Flight Model Characteristics

	1967-1968			1970-1971		
	Size ft^3	Weight lbs	Power watts	Size ft^3	Weight lbs	Power watts
Display Processor	0.43	19	65	0.20	12	40
Computer Memory	0.15	7	30	0.07	3	15
Pattern Memory	0.15	5	60	0.10	4	20
Control Logic	0.06	3	10	0.06	2	7
Buffering	0.14	10	50	0.08	6	30
EL Screen	0.25	20	20	0.25	15	20
EL Switching	0.30	12	150	0.20	7	65
Power Supplies	1.00	60	200	0.50	30	80
Manual Control	0.10	3	5	0.10	3	5
TOTALS	2.58	139	590 *	1.56	82	282 **

* Standby Power of 265 watts

** Standby Power of 190 watts

At this point in the program it can be concluded that:

- The EL crossed-grid display is a leading candidate for the ultimate solid-state display medium.
- The format versatility of an EL screen allows the use of the display formats not available with electromechanical devices.
- A multiformat display has data processing requirements which require further analysis and optimization.

BIBLIOGRAPHY

1. Raymond, R. G., Honeywell Memo 1821/RGR/53 dated 21 December 1966, EL-7 Logic Design.
2. Acker, J. R., Honeywell Memo 1821/JRA/63 dated 23 February 1967, Thin Film Brightness Control Parameters.
3. Wingert, J. W., Honeywell Memo dated 24 October 1966, Simulation Report: CRT Attitude Display Evaluation.
4. Omoth, R. E., Honeywell Memo 1821/REO/77 dated 13 April 1967, Computer Program for Electroluminescent Display Operation.
5. Paulson, D. C., Honeywell Report, "Computational Requirements of Solid State Displays," 1 March 1967.
6. Raymond, R. G., Honeywell Memo 1821/RGR/49 dated 13 January 1967, Size, Weight and Power Estimates for EL Display.

ACKNOWLEDGEMENT

The authors appreciate the efforts of the individuals associated with this project at Honeywell Inc. They particularly wish to express gratitude for the leadership and encouragement provided by project engineer, J. A. Miller.

67-552

-- NOTES --

No. 67-553

APPLICATION OF REDUNDANCY IN THE SATURN V GUIDANCE AND CONTROL SYSTEM

by

F. B. MOORE and J. B. WHITE
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-553

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.01, 8.04, 8.08

APPLICATION OF REDUNDANCY IN THE SATURN V GUIDANCE AND CONTROL SYSTEM

F. B. Moore and J. B. White
Guidance and Control Division, Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

The Saturn launch vehicle's guidance and control system is so complex that the reliability of a simplex system is not adequate to fulfill mission requirements. Thus, to achieve the desired reliability, redundancy encompassing a wide range of types and levels was employed. At one extreme, the lowest level, basic components (resistors, capacitors, relays, etc.) are employed in series, parallel, or quadruplex arrangements to insure continued system operation in the presence of possible failure conditions. At the other extreme, the highest level, complete subsystem duplication is provided so that a backup subsystem can be employed in case the primary system malfunctions. In between these two extremes, many other redundancy schemes and techniques are employed at various levels. Basic redundancy concepts are covered to gain insight into the advantages obtained with various techniques. Points and methods of application of these techniques are included. The theoretical gain in reliability resulting from redundancy is assessed and compared to a simplex system. Problems and limitations encountered in the practical application of redundancy are discussed as well as techniques verifying proper operation of the redundant channels. As background for the redundancy application discussion, a basic description of the guidance and control system is included.

Nomenclature

A	ratio of failures detected by current sensing to all failures in a duplex memory
F	number of units that have failed in a simplex system after time t
k	environmental adjustment factor
m	total number of trials in simulated sampling
N	number of remaining good elements in a simplex system after time t
N_d	number of duplex memory pairs in series
N_i	number of components of type i
N_o	number of components or elements comprising a simplex system
n	number of modules in a simplex computer
$P, \overline{P}$	probability of success and failure, respectively, of a redundant arrangement or system
P_a	actual but unknown system reliability
P_e	estimated reliability obtained through sampling
P_g	reliability gained by considering failures in opposite directions cancelling in a TMR digital arrangement
P_h, $P_{h'}$	probability that the h and h' converter, respectively, in a duplex power supply is good
$\overline{P}_{h\ low}$, $\overline{P}_{h'\ low}$	probability that the h and h' converter, respectively, in a duplex power supply fails low
$p_1, p_2, \cdots p_n$	probability that the events, $\xi_1, \xi_2, \cdots \xi_n$, respectively, will occur
$R, \overline{R}$	probability of success (or reliability) and failure, respectively, of a simplex unit
R_1, R_2	reliability of memory modules 1 and 2, respectively, of a duplex pair
R_a, R_b, R_c, R_d	probability of success of a simplex unit denoted by the subscript
$\overline{R}_{a1}, \overline{R}_{b1}, \overline{R}_{c1}$	probability that the a, b, or c unit, respectively, fails to a logical "1"
R_{a0}, R_{b0}, R_{c0}	probability that the a, b, or c unit, respectively, fails to a logical "0"
R_e, R_f	reliability of a power supply (or excitation source) and a simplex feedback amplifier, respectively
R_h, R_k	reliability of a simplex converter and the accelerometer encoder and signal conditioning circuitry, respectively
R_i, R_j	reliability of the logic and an actuator-servoamplifier channel, respectively
R_m, R_o	multiplexer and oscillator reliability, respectively

R_q, R_s	reliability of the subtract and limit check circuitry and a switch, respectively
R_t	reliability of platform sliprings, gimbal angle resolver, and two crossover detectors
R_v, R_w	reliability of a decision element (or voter) and a hydraulic supply, respectively
$R_{\dot{\varphi}}$, R_Ψ	reliability of an attitude rate command channel and an attitude command channel, respectively
S	possible states of an element
T, t	total mission time and operating time, respectively
t_{ik}	time at which the k^{th} failure of i^{th} type component occurs
U	unreliability or probability of system failure, expressed in terms of failures per million
U_r	unreliability of a redundant arrangement, expressed in failures per million
U_{ra}	unreliability of the redundant platform system through the orbital injection phase, expressed in failures per million
U_{rb}	unreliability of the redundant platform system during earth orbit and lunar injection phase, expressed in failures per million
U_{rab}	unreliability of the redundant platform system during all flight phases, expressed in failures per million
U_{rp}	unreliability of the redundant portion of an arrangement containing both redundancy and simplex units, expressed in terms of failures per million
U_s	unreliability of a simplex subsystem or system, expressed in failures per million
U_{sp}	unreliability of the simplex portion of an arrangement containing both redundancy and simplex units, expressed in terms of failures per million
V_a, V_b, V_c	decision element state denoted by the subscript
Z_c	confidence limit expressed in terms of standard deviations
λ, λ_i	unit failure rate and failure rate of the i^{th} component, respectively
ξ_1, ξ_2, --- ξ_n	independent events with probabilities p_1, p_2, --- p_n, respectively

Introduction

The development of the Saturn V launch vehicle system may be traced through successive developments of the Saturn I and Saturn IB vehicles, which consist of two propelled stages and an Instrument Unit. The first stage (S-I) of Saturn I consisted of eight engines with a combined thrust of 6.7×10^6 N (1.5 million lb); the second stage (S-IV) has six LH_2/LOX engines with a total thrust of 0.4×10^6 N (90,000 lb). A boilerplate of the Apollo spacecraft was flown with Saturn I. The first stage (S-IB) of Saturn IB has the same basic eight engine configuration as the Saturn I, but the engines have been modified to increase performance to a total thrust of 7.1×10^6 N (1.6 million lb). The second stage (S-IVB) of Saturn IB has one large LH_2/LOX engine with a thrust of 0.9×10^6 N (200,000 lb). The Instrument Unit in both vehicles provides guidance and control, vehicle sequencing, telemetry, and other instrumentation.

The Saturn IB system, whose maiden flight occurred early in 1966, bridges the gap between the Saturn I and Saturn V vehicles. This system consists of concepts and hardware developed for the Saturn I program and incorporates new ideas, techniques, and hardware required in the Saturn V system. It has the capability of orbiting the Apollo spacecraft.

In the Saturn V system, which is being developed to place a man on the moon, the second stage (S-IVB) of the Saturn IB vehicle moves up to become the third stage. Likewise, the Instrument Unit and the payload remain basically intact and make up the forward portion of the vehicle. The first stage (S-IC) consists of five newly developed engines; each has a thrust approximately equivalent to that of the total Saturn I first stage, and the total thrust is 33.5×10^6 N (7.5 million lb). The second stage (S-II) is being developed with five LH_2/LOX engines, each with a thrust equivalent to that used on the S-IVB stage; the total thrust is 4.5×10^6 N (1.0 million lb). The Instrument Unit of the Saturn V vehicle is basically equivalent to that of Saturn I and IB with slight modifications or equipment rearrangement to accommodate and facilitate the Apollo mission. The Saturn V guidance and control system discussed applies generally to the Saturn IB system as well.

The primary mission of the Apollo project is to place three astronauts in a lunar orbit, to land two of the astronauts on the moon's surface, and to safely return the crew to the earth's surface. The Saturn V launch vehicle is instrumental in the first phase of this operation for it is the vehicle system that will inject the spacecraft and its crew into the lunar trajectory. Since

so much is at stake in this project, both in terms of the lives of entire crews as well as the tremendous expense of such an undertaking, it is imperative that each mission be successfully completed. Considerable effort has been expended from the outset of the conceptual design phase to insure that the Saturn V launch vehicle is as reliable as today's technology permits. In many cases, the technology has been extended considerably to meet the stringent reliability requirements for these complex missions. In addition to the Apollo mission, it is expected that the Saturn V vehicle system will be required for other critical earth orbit and possibly interplanetary missions.

Major emphasis has been placed on attaining the highest reasonable reliability in the development of the flight-critical guidance and control system of the Saturn V launch vehicle. The emphasis on reliability has overshadowed other design considerations such as minimized weight, power consumption, and, to some extent, cost.

Historically, reliability improvement has been attacked through simplicity in concept, conservative design, high reliability component parts, and extensive testing programs and techniques. These basic principles have been extensively employed in the guidance and control system design. The number and type of functional units required to fulfill the prescribed mission have been kept to the absolute minimum. The hardware in the Saturn system is conservatively designed with flight-proven components and techniques being employed to the maximum extent. In spite of the conservatism and emphasis on simplicity employed in the basic system layout and detailed hardware design, the implemented system is still extremely complex, consisting of millions of component parts which must operate over extended periods of time. Therefore, redundancy is required to achieve the desired reliability.

Basic Redundancy Concepts

Within the past two decades, tremendous strides have been made in improving component part reliability. The transistor demonstrated a marked reliability improvement in comparison to the electronic tube; and, in more recent years, microminiaturization and integrated circuits have contributed significantly to electronic circuit reliability improvement. However, even with this advancement in basic technology, overall system reliability has not improved sufficiently to meet today's demand for the following reasons. First, the number of component parts in today's systems has increased significantly compared to those of a few years ago. Second, reliability requirements have increased considerably because of man-rated systems and the necessity of extended periods of operation. For these reasons, new techniques utilizing redundancy concepts have been developed. The concepts themselves are not new and were investigated by J. von Neumann and others; however, only recently have they been employed on such a large scale. The Saturn V guidance and control system represents the largest scale application of redundancy that exists in any present flight system.

The types of redundancy employed fall into the following categories: duplex, triple modular redundant (TMR), prime-reference-standby (PRS), quadruplex, and multiple parallel elements (MPE). Each approach is discussed to point out the reliability improvement obtained.

Three axioms of probability theory useful in the following derivations of reliability are as follows.

1. If p denotes the probability that an event will occur, then 1-p denotes the probability that the event will not occur.

2. If the events ξ_1, ξ_2, --- ξ_n are independent events with probabilities p_1, p_2, --- p_n, respectively, then the probability that all of the events should happen simultaneously when all are in question is the product of the probabilities

$$p = \prod_{i=1}^{n} p_i . \qquad (1)$$

3. If the probabilities of mutually exclusive events ξ_1, ξ_2, --- ξ_n are p_1, p_2, --- p_n, respectively, then the probability that any one of these events should happen when all are in question is the sum of the probabilities

$$p = \sum_{i=1}^{n} p_i . \qquad (2)$$

The reliability or probability of success of a single unit, whether a single component or a system, will be represented by R, and the reliability of the redundant arrangement by P. It is assumed that the equipment under discussion has been operated through a burn-in phase and does not have or has not reached the wearout phase. The reliability can therefore be conveniently expressed as a time dependent function. The expression relating reliability to time may be simply derived as follows.

Consider that N_o integral units, either single components or subsystems, comprise a system. Assume that each unit is functioning independently of the others and that the number of units which have failed at time t is F. Then, the number of good units (N) remaining after time (t) is

$$N = N_o - F . \qquad (3)$$

Assuming that the failure rate of the units is directly proportional to the number of good units results in

$$\frac{dF}{dt} = \lambda N \qquad (4)$$

where λ is the constant of proportionality and is commonly referred to as unit failure rate.

Substituting equation 3 into equation 4 results in

$$\frac{dF}{dt} = \lambda (N_o - F) . \qquad (5)$$

Solving this differential equation for F and evaluating the solution at t = 0 and F = 0 for the constant of integration yields

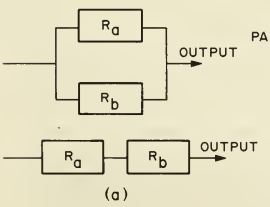
$$F = N_o (1 - e^{-\lambda t}) . \qquad (6)$$

If a unit is selected from the set, the probability that it has failed is, by definition, F/N_o; from axiom 1 the probability that it is good is $1 - F/N_o$ or from equation 4 is given by

$$R = e^{-\lambda t} \qquad (7)$$

The simplest and lowest level of redundancy utilized is that which duplicates a component part to prevent a system failure in the presence of a short or open of the component. With a component that tends to fail in the shorted mode, an additional component would be added in series; likewise, for a predominant open failure mode, a parallel component would be added. These arrangements are shown symbolically in Figure 1; the truth table represents the possible states of the units. The total number of combinations of states is derived from S^n, where S is the number of possible states and n is the number of units. In this arrangement, there are two states since each unit can either be good or bad, and the number of units is two, giving four possible combinations. If in the truth table a "0" is interpreted as a failure in the predominant mode and a "1" represents an operative unit, the same table applies to both the series and parallel combinations.

Truth tables, which are of primary importance in the design of logical systems, are useful in enumerating the possible combinations or states of a system and selecting the combinations which result in a system failure as well as indicating the assumptions and failure modes in each case. With a truth table and axioms 1, 2, and 3, the Boolean expression for system reliability can be readily derived. This technique will be used throughout to derive the reliability expressions.



TRUTH TABLE FOR SERIES OR PARALLEL COMPONENTS

R_a	R_b	SYSTEM STATUS
0	0	FAILED
0	1	OPERATIVE
1	0	OPERATIVE
1	1	OPERATIVE

(b)

Figure 1. Series and Parallel Configuration with Truth Table

The probability that the system is operative is given by

$$P = \bar{R}_a R_b + R_a \bar{R}_b + R_a R_b . \qquad (8)$$

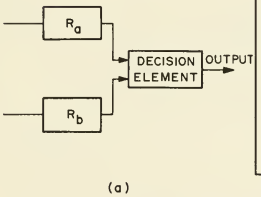
Assuming $R = R_a = R_b$, we obtain

$$P = (1-R) R + R (1-R) + R^2 = 2R - R^2. \qquad (9)$$

The reliability of the system as a function of time and unit failure rate is obtained by substituting equation 7 into equation 9 which results in

$$P = e^{-\lambda t} (2 - e^{-\lambda t}) . \qquad (10)$$

The duplex arrangement can also be employed at the module and subsystem level, where a single predominant failure mode cannot be assumed to exist. In this arrangement a decision element to determine which channel is operating correctly must be added. Consider a duplex arrangement, composed of identical units, and a decision element with the ability to determine which of the two units is good in case of a unit failure. This is shown symbolically in Figure 2; the truth table represents the possible states of the unit.



TRUTH TABLE FOR DUPLEX UNITS

R_a	R_b	DECISION ELEMENT STATE	SYSTEM STATUS
0	0	B	FAILED
0	0	A	FAILED
0	1	B	OPERATIVE
0	1	A	FAILED
1	0	B	FAILED
1	0	A	OPERATIVE
1	1	B	OPERATIVE
1	1	A	OPERATIVE

(b)

Figure 2. Duplex Configuration with Truth Table

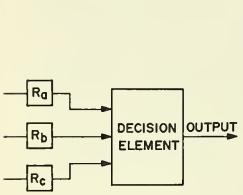
In the truth table, a "0" is interpreted as a failed unit and a "1" represents an operative unit. The A or B in the decision element state column indicates which element has been selected. It has been assumed that the decision element must select one element, but that both cannot be selected simultaneously. The logical conditions necessary for the system to be operative are $\bar{R}_a \cdot R_b \cdot V_b + R_a \cdot \bar{R}_b \cdot V_a + R_a \cdot R_b \cdot V_b + R_a \cdot R_b \cdot V_a$ where V_b and V_a indicate which unit has been selected. The reliability of the duplex system, when the reliability of the decision element is considered, is given by

$$P = R^2 + 2 (R - R^2) R_v \qquad (11)$$

where R_v is the decision element reliability. This equation reduces to that for the series or parallel cases (equation 10) if the reliability of the decision element is ignored; i.e., it has a reliability of one.

The duplex technique is one of the most desirable forms of redundancy, both in terms of simplicity and reliability improvement. However, the major disadvantage which limits its application considerably is the problem of determining the functional unit when a failure has occurred. The techniques used in the Saturn system to overcome this shortcoming are discussed later.

A triplex, or triple modular redundant (TMR), arrangement is shown in Figure 3. In this system the decision element, sometimes called voter, reacts to the majority inputs; consequently, only one failure can be tolerated.



(a)

TRUTH TABLE FOR TMR UNITS

Ra	Rb	Rc	SYSTEM STATUS
0	0	0	FAILED
0	0	1	FAILED
0	1	0	FAILED
0	1	1	OPERATIVE
1	0	0	FAILED
1	0	1	OPERATIVE
1	1	0	OPERATIVE
1	1	1	OPERATIVE

(b)

Figure 3. TMR Configuration with Truth Table

Four of the combinations result in system failure while the other four yield proper operation. The Boolean expression for proper operation is

$$P = \bar{R}_a R_b R_c + R_a \bar{R}_b R_c + R_a R_b \bar{R}_c + R_a R_b R_c \quad (12)$$

indicating that only one failure can be tolerated. Therefore, assuming identical units, the reliability of the system is given by

$$P = 3\,(1-R)\,R^2 + R^3 = 3R^2 - 2R^3. \quad (13)$$

The reliability of the decision element in a TMR arrangement may be considered in one of two ways. If three decision elements are used per trio, i.e., one for each element, the reliability of the voter may be lumped with that of the unit. The reliability of the unit then is decreased accordingly. If a single decision element is used for a trio, the result is a trio in series with a single element resulting in a reliability given by

$$P = (3R^2 - 2R^3)\,R_v \quad (14)$$

where R_v is the voter reliability. In either case, when the voter is assumed to be perfect, $R_v = 1$, the reliability of the system is given by equation 13.

Where TMR techniques are utilized in digital applications, advantage can be taken of the possibility of failures in opposite directions cancelling. For example, the second combination in the truth table (Fig. 3) would not have resulted in a system failure if R_a had failed to a logical "0" and R_b to a logical "1," or if R_a had failed to a logical "1" and R_b to a logical "0." This may be expressed in the form

$$\bar{R}_{a0} \cdot \bar{R}_{b1} \cdot R_c + \bar{R}_{a1} \cdot \bar{R}_{b0} \cdot R_c$$

where the second subscript indicates failure mode. Since this can occur in three such combinations, the Boolean expression for the reliability gained by opposite failures cancelling is

$$P_g = \bar{R}_{a0} \cdot \bar{R}_{b1} \cdot R_c + \bar{R}_{a1} \cdot \bar{R}_{b0} \cdot R_c + \bar{R}_{a0} \cdot R_b \cdot \bar{R}_{c1} + \bar{R}_{a1} \cdot R_b \cdot \bar{R}_{c0} + R_a \cdot \bar{R}_{b0} \cdot \bar{R}_{c1} + R_a \cdot \bar{R}_{b1} \cdot \bar{R}_{c0} \cdot \quad (15)$$

The probability of unit failure is the sum of the probabilities of component failures to a "0" state and to a "1" state; thus $\bar{R} = \bar{R}_o + \bar{R}_1$. Without investigating the details of a specific application there is no reason to suspect a failure to any particular state to be more prevalent than to the other state; consequently, $\bar{R}_o = 1/2\,\bar{R}$ and $\bar{R}_1 = 1/2\,\bar{R}$. This leads to the conclusion that $\bar{R}_o = 1/2\,(1-R)$ and $\bar{R}_1 = 1/2\,(1-R)$. Substituting these values into equation 15 yields the reliability gained from consideration of failures in opposite directions and is given by

$$P_g = 6\,[(R)\ 1/2\,(1-R)\ 1/2\,(1-R)]$$

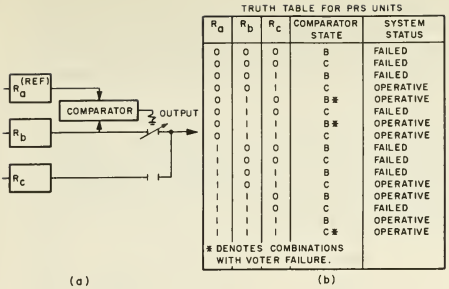
$$= \frac{3R}{2}\,[(1-2R+R^2)]. \quad (16)$$

The reliability of a TMR system when failures in opposite directions are considered is given by the sum of equations 13 and 16 yielding

$$P = (3R^2 - 2R^3) + \left(\frac{3R}{2} - 3R + 3/2\,R^3\right)$$

$$= 1/2\,(3R - R^3). \quad (17)$$

Another redundancy scheme is the primary-reference-standby (PRS) technique employing three channels that serve, as the name implies, three separate functions. In the normal unfailed condition, the primary channel B is functional in the system. Its output is compared to the reference A; and, in case of disagreement beyond an established level, the standby channel C is substituted for B. This scheme along with its truth table is shown in Figure 4.

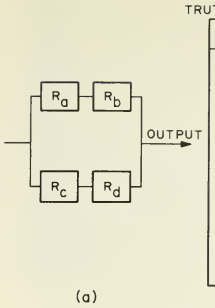


TRUTH TABLE FOR PRS UNITS

R_a	R_b	R_c	COMPARATOR STATE	SYSTEM STATUS
0	0	0	B	FAILED
0	0	0	C	FAILED
0	0	1	B	FAILED
0	0	1	C	OPERATIVE
0	1	0	B*	OPERATIVE
0	1	0	C	FAILED
0	1	1	B*	OPERATIVE
0	1	1	C	OPERATIVE
1	0	0	B	FAILED
1	0	0	C	FAILED
1	0	1	B	FAILED
1	0	1	C	OPERATIVE
1	1	0	B	OPERATIVE
1	1	0	C	FAILED
1	1	1	B	OPERATIVE
1	1	1	C*	OPERATIVE

* DENOTES COMBINATIONS WITH VOTER FAILURE.

Figure 4. PRS Configuration with Truth Table



TRUTH TABLE FOR QUADRUPLEX UNITS

R_a	R_b	R_c	R_d	SYSTEM STATUS
0	0	0	0	FAILED
0	0	0	1	FAILED
0	0	1	0	FAILED
0	0	1	1	FAILED
0	1	0	0	FAILED
0	1	0	1	OPERATIVE
0	1	1	0	OPERATIVE
0	1	1	1	OPERATIVE
1	0	0	0	FAILED
1	0	0	1	OPERATIVE
1	0	1	0	OPERATIVE
1	0	1	1	OPERATIVE
1	1	0	0	FAILED
1	1	0	1	OPERATIVE
1	1	1	0	OPERATIVE
1	1	1	1	OPERATIVE

Figure 5. Quadruplex Configuration with Truth Table

Again it has been assumed that the comparator has selected either B or C, but that it cannot select both simultaneously. The necessary logical conditions for this system to be operative are

$$\bar{R}_a \cdot \bar{R}_b \cdot R_c \cdot V_c + \bar{R}_a \cdot R_b \cdot \bar{R}_c \cdot V_b + \bar{R}_a \cdot R_b \cdot R_c \cdot V_b$$

$$+ \bar{R}_a \cdot R_b \cdot R_c \cdot V_c + R_a \cdot \bar{R}_b \cdot R_c \cdot V_c + R_a \cdot R_b \cdot \bar{R}_c \cdot V_b$$

$$+ R_a \cdot R_b \cdot R_c \cdot V_b + R_a \cdot R_b \cdot R_c \cdot V_c$$

where V_b and V_c indicate which element has been selected. When the units are assumed to be identical, the reliability of the system is given by

$$P = (R^3 - R^2)\ (1 - 2R_v) + R \qquad (18)$$

where R_v is the comparator reliability. If the comparator is assumed to have a reliability of one, equation 18 reduces to

$$P = R\,(1 + R - R^2)\,. \qquad (19)$$

The PRS technique has a major disadvantage in that is is more susceptible to transients or intermittents than the other schemes. Consequently, if a transient causes the comparator to switch to the standby unit, means should be available to switch back to the original unit with its reference; otherwise all the advantages of the redundant system have been lost from that point on. As discussed later, the switchback technique is employed in some PRS portions of the Saturn system but not in others.

The next technique to be considered is the quadruplex arrangement shown with its truth table in Figure 5. Since the arrangement has four units, 2^4 combinations are possible. In Figure 5, assume that only one failure in each or in both branches can be tolerated, and two failures in any one branch will result in a system malfunction.

Inspection of the truth table for the quadruplex arrangement reveals that the system reliability may be obtained by

$$P = 1 - (1-R)^4 - 4(1-R)^3 R - 2(1-R)^2 R^2$$

$$= R^2\,(4 - 4R + R^2)\,. \qquad (20)$$

The quadruplex arrangement is most useful when applied at the component level, i.e., to resistors, capacitors, diodes, valves, relays, etc., where the component does not have a single predominant failure mode. In applications where a single failure mode exists, two components in series or parallel would be employed in preference to the quadruplex arrangement.

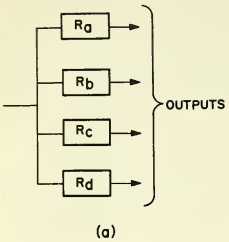
An inherent redundancy exists in some subsystems because of certain features of the overall system configuration dictated by other subsystems. In such cases the subsystem may continue to operate either with no degradation or with an acceptable degradation of performance in the presence of one or more failed elements. An example of such a situation exists in the Saturn guidance and control system because of the required clustering of engines to provide the necessary vehicle thrust. Since four engines are gimbaled to maintain vehicle control, the failure of one of the four control channels in each plane does not cause a system failure. The subsystem can be treated as one having four parallel elements, with the failure of any one element being permissible. This arrangement is referred to as multiple parallel elements (MPE). The applicable schematic and truth table are shown in Figure 6. Five combinations in Figure 6 result in continued successful operations. The resulting expression for proper operation is

$$P = \bar{R}_a R_b R_c R_d + R_a \bar{R}_b R_c R_d + R_a R_b \bar{R}_c R_d$$

$$+ R_a R_b R_c \bar{R}_d + R_a R_b R_c R_d\,.$$

Again, assuming identical units results in

$$P = 4(1-R)\,R^3 + R^4 = 4R^3 - 3R^4\,. \qquad (21)$$



(a)

TRUTH TABLE FOR MPE UNITS

Ra	Rb	Rc	Rd	SYSTEM STATUS
0	0	0	0	FAILED
0	0	0	1	FAILED
0	0	1	0	FAILED
0	0	1	1	FAILED
0	1	0	0	FAILED
0	1	0	1	FAILED
0	1	1	0	FAILED
0	1	1	1	OPERATIVE
1	0	0	0	FAILED
1	0	0	1	FAILED
1	0	1	0	FAILED
1	0	1	1	OPERATIVE
1	1	0	0	FAILED
1	1	0	1	OPERATIVE
1	1	1	0	OPERATIVE
1	1	1	1	OPERATIVE

(b)

Figure 6. MPE Configuration with Truth Table

Any of the redundant arrangements may be cascaded and the total system reliability may be found from axiom 2. For example, a system composed of two duplex subsystems similar to those in Figure 2 would have a reliability given by

$$P = (2R-R^2)^2. \qquad (22)$$

Similarly, a system composed of a duplex subsystem and a TMR subsystem would have a reliability given by

$$P = (2R-R^2)\ (3R^2-2R^3). \qquad (23)$$

To summarize, Table 1 shows the reliability expression for each scheme discussed in order of reliability preference. However, practical limitations usually determine the choice of schemes.

Table 1. Redundancy Schemes

Scheme	Reliability Expression	Assumptions
Duplex	$2R-R^2$	Proper decision element can be determined.
TMR	$1/2\,(3R-R^3)$	Failures in opposite directions can cancel.
PRS	$R\,(1+R-R^2)$	Reference and normally used unit do not fail to the same state simultaneously.
Quadruplex	$R^2\,(4-4R+R^2)$	Limited generally to component part application.
MPE	$4R^3-3R^4$	4 elements
Simplex	R	

In applying the theory to the assessment of the reliability of a complex system, it is sometimes more convenient to express reliability equations in terms of unreliability which can be derived from unit failure rates. The unreliability of each component, subsystem, or system is then expressed as a number of failures per unit of time, permitting easier separation or combination of the associated numbers without resorting to the manipulation of numbers involving a series of "nines."

Since R in the reliability expressions may be replaced by $1-\bar{R}$ where $\bar{R}$ is the probability of subsystem failure, the reliability of a redundant unit may be expressed in terms of the probability of failure of the single nonredundant unit. The result for each type of redundancy is as follows.

Duplex	$P = 1-\bar{R}^2$
TMR	$P = 1/2\ [2-3\bar{R}^2+\bar{R}^3]$
PRS	$P = 1-2\bar{R}^2+\bar{R}^3$
Quadruplex	$P = 1-2\bar{R}^2+\bar{R}^4$
MPE	$P = 1-6\bar{R}^2+8\bar{R}^3-3\bar{R}^4$

(24)

Further, $R = e^{-\lambda t}$ and $\bar{R} = 1-e^{-\lambda t} \approx 1-(1-\lambda t+\text{---}) \approx \lambda t$, for very small λt. Since in equations 24, $\bar{R}$ is also very small, terms higher than the second order may be ignored. If the higher order terms are ignored, the approximations for redundant system unreliability expressed in terms of component failure rates and operating time are

Duplex	$\bar{P} \approx (\lambda t)^2$
TMR	$\bar{P} \approx 3/2\,(\lambda t)^2$
PRS	$\bar{P} \approx 2\,(\lambda t)^2$
Quadruplex	$\bar{P} \approx 2\,(\lambda t)^2$
MPE	$\bar{P} \approx 6\,(\lambda t)^2$

(25)

From equations 25, the ordering of the system in rank of reliability becomes obvious.

Figure 7 is a graphical comparison of the reliability of the simplex, duplex, TMR, PRS, MPE, and quadruplex schemes as a function of unit failure rate and time, where $R = e^{-\lambda t}$ has been substituted into the equations previously derived. In the case of the TMR arrangement, failures in opposite directions cancelling were assumed; for the PRS arrangement, it was assumed that the reference unit and the unit to which it is normally compared do not fail simultaneously to a state which cannot be detected by the comparator. The figure further substantiates the relative desirability of each scheme. The fact that a portion of the reliability curve of the quadruplex and MPE scheme falls below that of a simplex system is not significant because this occurs at a reliability far below that which would be permissible in a practical application. It is interesting to note that in the region above 0.9, the reliability of the quadruplex and PRS schemes is practically identical (equations 24 and 25).

Figure 8 further demonstrates the merits of redundant systems compared to a simplex system and indicates quantitatively what can be gained through the

various techniques. For convenience, unreliability in terms of failures per million is shown for both the simplex and redundant systems. In the reliability assessment and comparisons appearing in the following sections, the quantities are expressed in these terms.

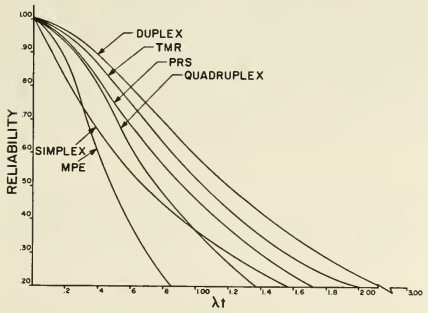


Figure 7. Reliability Versus λt for Various Redundancy Schemes

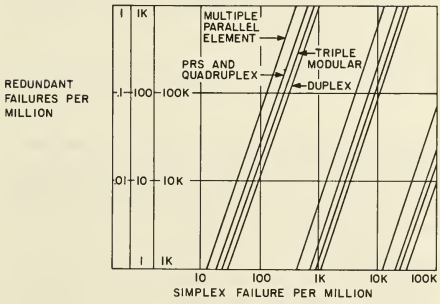


Figure 8. Redundant Failure per Million Versus Simplex Failures per Million for Various Redundancy Schemes

Guidance and Control System Description

The Saturn navigation, guidance, and control system is completely self-contained within the vehicle and utilizes onboard inertial sensors, computation, and control to direct the vehicle according to the desired path and end conditions. A digital command system is available as a part of the onboard astrionics system, but is not planned for use in the primary mode.

The navigation function is accomplished through the use of acceleration measurements provided by accelerometers mounted on the space-direction-fixed stable element of the stabilized platform. The resulting information is processed within the onboard digital computation system. The acceleration information is integrated to obtain vehicle velocity and position information. The current measured position information is used to continuously calculate and combine the gravitational effects with the measured data to obtain space-fixed vehicle velocity and position.

The guidance function, which is the computation of the necessary maneuvers to satisfactorily reach the specified end conditions, is accomplished within the onboard digital computer system. To give the desired result, the implemented guidance equations must take into account various mission and vehicle constraints, one of the most significant of which is that of propellant consumption optimization. The equations programed into the onboard digital computer system represent a path adaptive guidance scheme, termed the iterative guidance mode (IGM), which fulfills the optimization requirements and the guidance requirements for insertion both into earth orbit and injection into the lunar trajectory. The specific results of the guidance computation are as follows.

1. Instantaneous required thrust direction expressed as three Euler angles.

2. Required time of engine cutoff to achieve the specified orbital conditions.

3. Required time of engine ignition to leave earth orbit.

4. Required time of second cutoff to satisfy the lunar trajectory end conditions.

The required angular directions resulting from the guidance calculations are applied to the vehicle through the control system. In addition to responding to the commands of the guidance system, the control system must maintain stabilization of the vehicle attitude in the presence of various vehicle propellant sloshing, structural bending, and load constraints. The elements of the control system required to accomplish this task can be divided into three specific functional areas: sensing of vehicle state information, computation, and vehicle torquing. In the Saturn V system, the vehicle state information required is that of attitude and rate. (On the Saturn I and IB vehicles, additional information obtained through vehicle-fixed lateral accelerometers was required to obtain structural load relief.) The attitude information is obtained from resolvers mounted on the stabilized platform gimbals. The information on actual vehicle orientation from the resolvers is compared in the onboard digital computer system with the desired orientation determined from the guidance calculations, resulting in the desired attitude control commands. The three-axis attitude rate information required to accomplish vehicle stabilization is obtained from vehicle-fixed rate gyros.

The control "computation" consists of the gain modification, filtering, mixing input attitude and rate information, and shaping of this information to provide vehicle stabilization in the presence of structural bending, propellant sloshing, and other dynamic

characteristics. Routing of the control signals to the proper end element to develop the desired vehicle controlling torques is also part of this function.

Two methods are used to develop the control torques in the Saturn V vehicle. Positioning of the primary propulsion engines by hydraulic actuators is used to control pitch and yaw on each of the three stages. In addition, control about the roll axis is obtained on the first two multiengine stages by the proper differential positioning of the gimbaled engines. Roll control on the single-engine third stage, and control of this stage about all three axes during coasting phases, is accomplished by an array of fixed direction thrusters. Pulses of thrust from these low thrust devices are commanded by the control electronics to provide corrective control torques about the appropriate vehicle axes.

The basic elements of the navigation, guidance, and control system are shown in block diagram form in Figure 9, which indicates the primary form of redundancy employed in each element. For a more detailed description, the system is broken down into the digital computer subsystem, the stabilized platform subsystem, and the control subsystem. Each of these subsystems encompasses a number of hardware elements, with many performing a variety of functions in the overall system.

The major systems are broken down in some instances to the "black box" level and in others to a specific functional level, depending on which breakdown is more convenient and appropriate to illustrate the application of redundancy. Although no attempt is made to describe in detail the total application of redundancy, examples of the different types are cited and described in each subsystem. Where available, reliability numbers are shown for the various modules in the subsystems as well as for the total subsystems. The theoretical gain in reliability through redundancy is also shown in each case. Since the reliability assessments of various elements were conducted by different groups, the numbers may not be universally compatible. However, some adjustment of the failure rates has been effected where obvious discrepancies existed between the numbers set forth in the various references. In spite of these adjustments, caution should be exercised in using the reliability numbers presented, even though the numbers do indicate in gross terms the relative reliability of the various elements and subsystems. The prime intent is not to provide an accurate and intensive reliability analysis, but rather to illustrate the benefits of the various redundancy techniques employed.

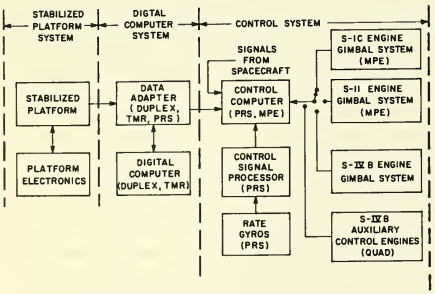


Figure 9. Saturn V Guidance and Control System

The simplified equations previously developed are used where possible. In many instances, the simplifying assumptions made in the development of those equations do not apply; therefore, specific equations that apply to the particular situation must be developed.

For convenience, the module, subsystem, and system assessments are expressed in terms of unreliability. Through this approach, the relative contributions of the various elements can be more easily portrayed. Additionally, with the simplifying assumptions made, the unreliability numbers of the various subelements can be added directly to obtain the total unreliability.

As previously shown for highly reliable systems,

$$R \approx \lambda t.$$

This approximation can be made with an error less than $(\lambda t)^2/2$.

In component or system operation in a particular application, a degradation factor to account for the effect of the particular environment must be considered. This is generally called the environmental adjustment factor, designated by k. Therefore $\bar{R} \approx k\lambda t$. The unreliability numbers are expressed as $U = k\lambda t \times 10^6$ indicating the number of failures per million flights. Note that the term "failures" as expressed here is intended to designate component or system malfunctions or out-of-tolerance operation in a million flights; it does not indicate the number of vehicle or mission losses in a million flights. To obtain the latter, which is not covered in this analysis, the failure modes of the various elements and the effects of those failures on the vehicle behavior would have to be additionally considered. Table II shows the k-factors for the various stages and the phase times used in deriving the unreliability numbers.

Table II. Phase Times and k-Factor for Various Stages

Flight Phase	k-factor				Phase time (hrs)
	Location of Component				
	S-IC Stage	S-II Stage	S-IVB	IU	
S-IC Burn	1500	900	700	115	0.041
S-II Burn	-	900	700	45	0.106
First S-IVB Burn	-	-	700	45	0.046
Orbital Coast	-	-	10	1	4.5
Second S-IVB Burn	-	-	700	45	0.087
Translunar Coast	-	-	10	1	2.0

Digital Computer System

The digital computer system developed for the Saturn V vehicle consists of two basic units, a launch vehicle digital computer (LVDC) and a launch vehicle data adapter (LVDA). The LVDC is the basic computing element in the vehicle with the capability of performing arithmetic operations such as add, subtract, multiply, and divide; it provides the intelligence for making logical choices. The LVDA is essentially the LVDC input/output unit and all signals to and from the LVDC are processed in this unit. In addition, it performs certain simple computational and logical operations on data. The computer system is instrumental in all three phases of operation for the Saturn V vehicle; i.e., it plays a major role in the automatic checkout of the vehicle before launch, solves the guidance equations, provides attitude correction signals and vehicle sequences during the boost phase, and assists in vehicle checkout during the orbital coast phase.

The LVDC is a serial, fixed-point, stored program, general purpose machine with a basic clock of 2.048 MHz. Four clocks comprise a bit time and 14 bits a phase time. The machine is organized to operate around three phases or cycles. For example, data may be read from memory during one phase or cycle and operated upon during the next two cycles. Data words 28 bits in length (25 magnitude bits, 1 sign, and 2 parity bits) are used in computation. The memory, which contains from one to eight random-access magnetic core modules each consisting of 4096 data words, is arranged in such a manner that one data word or two instructions (each instruction contains a parity bit) may occupy one 28-bit memory word. Special algorithms have been developed and implemented for multiplication and division; multiplication is done four bits at a time and division is done two bits at a time. The system utilizes microminiature circuitry where power and accuracy requirements permit. Where microminiaturization cannot be employed, conventional discrete components are used.

During flight, the digital computer system inputs are (1) platform accelerometer outputs, (2) platform gimbal angles representing vehicle attitude, (3) discrete inputs indicating vehicle functions such as liftoff, first stage cutoff, separation, second stage ignition, second stage cutoff, and engine out, and (4) command receiver signals allowing memory alteration and ground control.

During flight, the digital computer system outputs are (1) steering or attitude correction commands, (2) discrete outputs commanding vehicle sequencing such as cutoff and separation, and (3) telemetry data words, 40 bits each, at a maximum rate of 240 per second for monitoring trajectory parameters and computer system operation.

Because of the critical functions performed by this system, every effort has been made to make it as reliable as possible. Many forms of redundancy have been incorporated into the system, which utilizes quadruplex components and circuits, and duplex, TMR, and PRS techniques as well as overall system backups. The system represents one of the largest scale applications of redundancy employed to date. The LVDC and LVDA form a complex system containing more than 95,000 equivalent electronic components. Of this number, less than one half of one percent are employed in such a manner that a single component failure would result in a system failure.

Figure 10 shows a simplified block diagram of the LVDC and indicates the redundancy techniques employed in that unit, with the corresponding unreliability indicated in each block. The fact that the TMR timing and logic depicted in Figure 10 is very much simplified is borne out when the TMR organization of the LVDC is considered in any detail. For example, since the TMR logic of the machine is considered to consist of seven functional modules, in the idealized case, it would be expected that 21 voters would be employed in the machine. However, because of the various feedback paths and the fact that each module has several output signals feeding various other modules, the idealized model cannot be employed accurately. For example, instead of 21 voters being employed in the LVDC timing and logic, approximately 155 signals are voted on, giving a total of 395 voters. The LVDA employs 237 voters in its TMR logic.

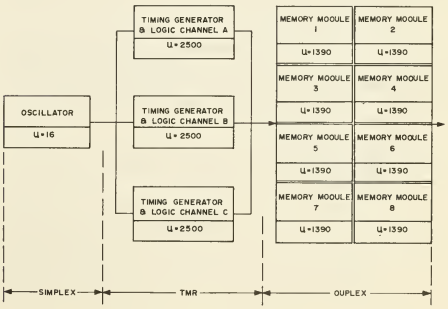


Figure 10. Block Diagram of the Launch Vehicle Digital Computer

67-553

Because of the relative simplicity of the basic 2.048 MHz oscillator (it contains only five electronic components) and the technical problems inherent in synchronizing multioscillators, a simplex oscillator system is employed in the LVDC. The output of the basic oscillator is used to form the necessary phasing and clock signals in the timing generator. Each channel of the TMR logic contains its own timing generator; consequently, a failure of the timing generator results in a failure of that channel. The memory system, expandable in modules of 4096 words, 28 bits in length, up to eight memory modules, is employed either in a duplex or simplex manner depending upon the criticality of the program being run. For instance, prelaunch programs are simplexed while flight routines are duplexed. From Figure 10, it is evident that the reliability of the LVDC may be approximated by

$$P = (R_o)\ (R_i)\ (R_m) \tag{26}$$

where R_o is the reliability of the simplex oscillator, R_i is the reliability of the combined TMR timing generator and logic, and R_m is the reliability of the duplex memories. The methods determining the reliability for each of these will now be considered.

The number of component parts in the system and their failure rate, the Saturn V mission time, and environmental conditions determine the unreliability of the oscillator which is $U_s = 16$.

The reliability for the timing generator and logic cannot be determined so simply for reasons indicated previously. Any attempt to accurately compute analytically the reliability of the timing generator and the complex logic it feeds, without making a great number of simplifying assumptions, would lead to a mathematical expression containing literally thousands of terms. Therefore, a method employing the Monte Carlo technique, which is basically a technique of simulated sampling, has been devised so that the reliability may be approximated.

Although the Monte Carlo technique is general and has been applied in many other fields, it represents a rather unique application in this particular field. Thus, a brief description of the basic procedures using this technique is in order; the evaluation procedures consist basically of three phases:

1. With simulation techniques, generate a set of failed components

2. Locate the computer subsystems containing the failed components

3. Trace the simulated failures through the logic to determine their consequence.

The first step consists of generating, by a random process, a set of failed components. If an exponential distribution of time to failure is assumed for a component, the probability of failure for that component is

$$\bar{R} = 1 - \epsilon^{-\lambda t} \tag{27}$$

where t is time and λ is the failure rate of the component. When the design contains N components of type i, the probability of failure becomes

$$\bar{R} = 1 - \epsilon^{-N_i \lambda_i t} \tag{28}$$

Solving equation 28 for t yields

$$t_{ik} = -\frac{\ln\ (1-\bar{R})}{N_i \lambda_i} \tag{29}$$

where t_{ik} is the time at which the k^{th} failure of component type i occurs. In each trial a random number between "0" and "1" is chosen and substituted for $\bar{R}$, and equation 29 is evaluated. The result is the time, t_{ik}, at which the first failure of component type i occurs. Then t_{ik} is compared to mission time T; if $t_{ik} < T$, a failure is recorded, N_i is reduced by one, and the process is repeated. As each t_{ik} is calculated, it is added to the sum of the previous t_{ik}'s and the new total is compared to mission time. The process is completed when the summation of the failure times exceeds the mission time, i.e., $\Sigma t_{ik} > T$.

Each of the i^{th} components in the system is assigned a number. The system's functional component that failed at time t_{ik} is determined by multiplying the random number generated by the total number of i components in the group; i.e., the random number chosen gives both time of failure and component that failed. This process is repeated for each component type in the system.

The second step consists of locating, within the logic framework of the machine, the component parts that failed. The third and final step consists of tracing the effects of the failed components, in the sequence in which they occur, upon the TMR logic. If in time T, the total combination of failures did not result in a system failure, a successful trial resulted. After many trials, the reliability of the system is then determined from

$$P = \frac{\text{number of successful trials}}{\text{total number of trials}}$$

The unreliability of the LVDC timing generator and logic using simulated sampling is $U_r = 10$. Approximately 20,000 "games" were played to determine this value. The confidence interval, which can be associated with this estimate as a function of the number of trials, is determined by

$$P_a = P_e \pm Z_c \sqrt{\frac{P_e\ (1 - P_e)}{m}} \tag{30}$$

where P_a is the actual but unknown reliability, P_e is the estimated reliability obtained from simulated sampling, Z_c is the confidence limit expressed in terms of standard deviations, and m is the total number of trials. From this, there is 90 percent confidence that $0 < U_r < 50$.

It is of interest to apply the simplified analytical technique derived earlier and to compare these results with those obtained from the Monte Carlo method. From the number of component parts in a simplex system and their failure rates and a Saturn V mission, the unreliability of a simplex computer timing and logic has been determined by Monte Carlo to be $U_s = 2500$.

A voter for a logic module adds approximately 25 percent to the number of component parts of that module; therefore, a simplex machine with enough component parts necessary for voters for one channel would have 25 percent more component parts than a simplex machine and would have a reliability given by

$$R = e^{-1.25n\lambda t}$$

where $n\lambda t = \ln (1-2500 \times 10^{-6})$. The unreliability of a simplex channel with voters then is $U_s = 3120$. If a simplex machine is divided into n modules, each of which has a reliability of $R^{1/n}$, and triplicated, the reliability of one trio as given by equation 17 is

$$P = 1/2 \left[3R^{1/n} - R^{3/n}\right]. \qquad (31)$$

Now, the reliability of a TMR machine consisting of n sets of triplicated modules is given by

$$P = 1/2\left[\left(3R^{1/n} - R^{3/n}\right)\right]^n . \qquad (32)$$

For the LVDC, since a simplex machine may be considered to have been divided in seven equivalent parts, n = 7, and R for each of the elements is 0.996880 as previously derived, the unreliability for the entire TMR logic is

$$U_r = 2 .$$

Since in the ideal case it was assumed that the seven logic modules have equal reliabilities and that the logic was organized in such a manner to utilize 21 voters (neither of which is true in practice), it is expected that the ideal case would result in higher reliability than that obtained through simulated sampling. The more accurate result derived by Monte Carlo techniques for the LVDC is U = 10, which is used in the subsequent assessment.

The reliability of the LVDC toroidal core memory system may be found directly from component part count and failure rates using analytical means. Since a major problem in duplex systems often is failure detection mechanisms, it is of interest to note the type of failure detection employed in the LVDC memory system. The memory has two types of failure detection circuitry: odd parity checking and half select current monitoring. It is felt that parity checking will detect major failures in the sense amplifiers, cores, inhibit drivers, memory buffer registers, and variable strobe gate; while half select current monitoring will indicate major failures in the voltage and current drivers, decoupling circuitry, memory timing, and connection circuitry. Errors not determined by current checking, however, may be detected by parity checking.

The reliability of the memory system may be found directly from the relationship

$$P = \left\{R_1 + R_2\,(1-R_1)\,[A + (1-A)\,(0.5)]\right\}^{N_d} \qquad (33)$$

where R_1 is the reliability of memory module 1 of a duplex pair, R_2 is the reliability of memory module 2 of the pair, A is ratio of failures detected by current sensing to all failures, (1-A) is ratio of failures not detected by current sensing to all failures, and N_d is the number of duplex pairs operating in series. Equation 33 infers that memory 1 is good or that memory 1 fails but 2 is good and that the failure is detected by the current sensing circuits or, if it is not detected, there is a 50/50 chance that it will be picked up with parity checking.

The reliability of a single memory module found from part count, generic failure rates, and Saturn V mission operating conditions is R = 0.998610; and from engineering design analysis the chances of a nondetectable failure (1-A) is 0.073. From equation 33, the unreliability of an eight memory module configuration with a storage capacity of 16,000 duplexed words is $U_r = 226$.

In summary, the unreliability of the LVDC for the Saturn V mission is the sum of the unreliabilities of the simplex oscillator, the TMR timing and logic, and the four duplexed memory modules, i.e., $U_r = 16 + 10 + 226 = 252$.

The reliability of the LVDA is not as straightforward as the LVDC because many varied functions entangled with the other systems, primarily the LVDC and platform, are performed in the LVDA. For example, the LVDA power supplies are required for operating the LVDC, processing vehicle attitude and velocity information, and issuing attitude correction commands. Parts of the LVDA TMR logic are time shared and are required with various critical vehicle functions. The LVDA utilizes various types of redundancy techniques: duplex, TMR, PRS, as well as system backup. However, only isolated types such as the LVDA power supplies and the digital-to-analog converter subsystem are discussed. All of these functions are flight critical. The reliability of the logic portion of the LVDA is found similarly to that for the LVDC. The TMR logic of the LVDA has an accessed unreliability of $U_r = 10$ for the Saturn V mission.

Six power supplies in the LVDA, which supply dc power to the LVDC as well as the LVDA, are all duplexed. Figure 11 shows a typical power supply. The dc to dc power converters are tied together through isolation diodes. Should a converter fail low, the other converter

picks up the load. It is imperative that a converter not fail high because the diodes isolation between the two units would be worthless. The feedback amplifiers used with each converter are duplexed to minimize the probability of this happening. One duplex system is functional provided the following condition is fulfilled.

$$P = \overline{P}_{h(low)} \cdot P_{h'} + P_h \cdot \overline{P}_{h'(low)} + P_h \cdot P_{h'}, \quad (34)$$

i.e., the output of converter h can be low and h' can be correct, or the output of converter h' can be low and h can be correct, or both outputs can be correct. Under the assumption that the chances of a feedback amplifier failing low are equal to those of it failing high (this is a valid and accurate assumption in this case), the expression for a simplex power supply (one converter and two feedback amplifiers) failing low is

$$\overline{P}_{h(low)} = \overline{P}_{h'(low)} = 1 - R_h R_f^2 - 2 R_h R_f (1-R_f) - R_h (1-R_f)^2 \quad (35)$$

where

$R_h R_f^2$ - probability that the converter and both amplifiers are good. (The output of the converter is correct.)

$2R_h R_f (1-R_f)$ - probability that the converter is good, one amplifier is good, and one has failed low. (The output of the converter is correct.)

$R_h (1-R_f)^2$ - probability that the converter is good and that both amplifiers have failed low. (The output of the converter is therefore high.)

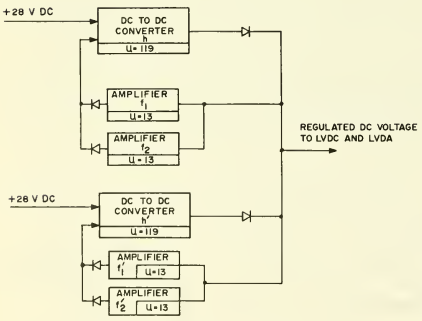


Figure 11. Typical LVDA Duplex Power Supply

Now, the probability that the output of a simplex power supply is correct is given by the first two terms and is

$$P_h = P_{h'} = R_h R_f^2 + 2R_h R_f (1-R_f). \quad (36)$$

Substituting equations 35 and 36 into equation 34 and simplying yields

$$P = [R_h R_f (2-R_f)] \; [2(1-R_h) + R_h R_f (2-R_f)]. \quad (37)$$

From generic failure rates, it has been determined that $R_f = 0.999987$ and $R_h = 0.999881$ yielding $U_r \approx 0$ for a duplex supply. In comparison, the reliability of a completely simplex supply, i.e., one converter and one feedback amplifier, is $R = R_h R_f$ which has an unreliability of $U_s = 132$.

Six supplies are used in the LVDA system; four have an unreliability of $U_r \approx 0$, and two which do not have isolation diodes because of high current requirements have an unreliability of $U_r = 5$. The unreliability of the complete LVDA duplexed power supply system is $U_r \approx 4(0) + 2(5) \approx 10$.

A block diagram of the LVDA digital-to-analog attitude correction conversion system is shown in Figure 12. The system accepts the attitude correction commands from the LVDC and converts them to an analog form which is compatible with the control computer. It therefore plays a vital and critical function in the guidance and control of the vehicle. The reliability scheme employed is basically a PRS system with a reference channel being compared with that which is normally active. Two comparators are used in the system; one is an accurate fine comparator while the other is a coarse comparator and compares the outputs from the sample and hold devices and the output amplifiers. The block diagram of Figure 12 can be further simplified to the PRS redundancy system shown in Figure 13. If this is done, the prime, the reference, and the standby units consist essentially of the nine-bit register, the ladder network, sample and hold circuits, and two amplifiers. The voter then consists of the fine and coarse comparator. A single failure in the channel select switch results in a loss of redundancy. (Although in many cases, multiple failures can be tolerated, particularly in the various subsystems within the vehicle, the basic ground rule used for subsystem design was toleration of one failure.) The reliability of the system (for all three axes) may be approximated by the expression derived earlier for this type of redundancy (equation 18) but must be modified to take into account the single failure mode of the switch. The approximate reliability is given

$$P = [(R^3 - R^2)(1 - 2R_v) + R] R_s \quad (38)$$

where R is the reliability of a channel, R_v is the reliability of the voter or comparator, and R_s is the reliability of the switch. It has been estimated that for a Saturn V mission, $R = 0.999648$, $R_v = 0.999956$, and

$R_s = 0.999912$ resulting in an overall digital-to-analog converter unreliability of $U_r = 89$. In comparison, the unreliability of a simplex system is $U_s = 352$. The redundancy has consequently resulted in a decrease in unreliability by a factor of 3.94.

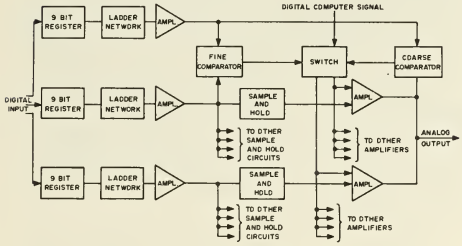


Figure 12. LVDA Digital-To-Analog Attitude Correction Conversion

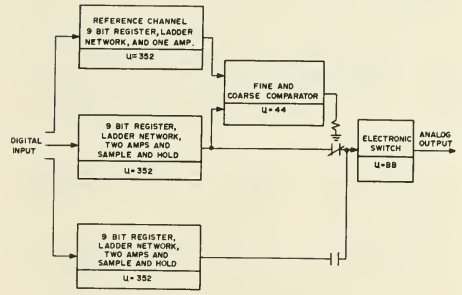


Figure 13. LVDA Digital-to-Analog Attitude Correction Conversion

Because of the nature of digital systems, intermittent failures are much more predominant than hard or solid failures. Therefore, the ability to switch from the standby unit back to the prime-reference system has been incorporated in the converter and the LVDC system. Since types of failures were not considered in the analysis, the reliability estimate is pessimistic from this standpoint.

In the LVDA, the reliability accessment has dealt with isolated examples, mainly the power supply and the digital-to-analog converter. Two other examples, that of processing attitude and acceleration inputs, are covered later. Table III summarizes the reliability of each major subsystem of the two units, both for the simplex and redundant case. Also shown is the ratio of the probability of failure of a simplex unit to that of a redundant unit. This factor indicates to some degree what has been gained through redundancy.

Table III. Summary of Unreliability of Simplex and Redundant LVDC and LVDA

Element	Simplex Unreliability U_s	Redundant Unreliability U_r	$\frac{U_s}{U_r}$
LVDC			
Logic	2,500	10	250
Memory (8mm duplex)	5,960	226	26.4
Oscillator	16		
Total LVDC	8,476	252	33.6
LVDA			
Power Supply	792	10	79.2
Input/Output	800	10	80.0
Logic	2,430	10	243
Total LVDA	4,022	30	134
Total Computer System	12,498	282	44.3

Stabilized Platform System

The stabilized platform is the basic reference for the Saturn navigation, guidance, and control systems. The system provides a space-direction-fixed coordinate reference frame which serves as a reference for the vehicle's attitude. The stabilized element serves as a base for three mutually-orthogonal accelerometers which provide the information from which translational velocity and position of the vehicle are derived. The stabilized platform system consists of an inertial platform, the associated electronics for internal stabilization and processing of output information, an electrical power supply, and a nitrogen gas supply.

The ST124-M platform used in the Saturn V system is a three-gimbal device which allows unlimited rotation of the vehicle about the pitch and roll axes. Rotation about the vehicle yaw axis (referenced to launch position) is limited to ±60 degrees, which is adequate to accomplish the Apollo mission. To accommodate missions requiring unlimited gimbal freedom about all three axes, the capability of incorporating a fourth gimbal has been designed into the system. On the three-gimbal platform, the order of the gimbals from the stabilized table outward is pitch, yaw, and roll, referenced to the vehicle position at liftoff. Dual-speed resolvers used as angular encoders on the gimbal pivots provide information from which the vehicle attitude is derived. Three single-degree-of-freedom gyroscopes provide the reference for the stable table on which the three pendulous integrating gyro accelerometers are mounted. Signal generators on the output axes of the reference gyroscopes derive electrical signals proportional to disturbance torques about the mutually perpendicular axes. These signals are amplified and shaped in the associated electronics and used to drive servotorque motors which maintain the inertial gimbal space-direction-fixed.

The inertial element of the reference gyroscope is a synchonous hysteresis gyro wheel having an angular momentum of 2.6×10^6 g cm^2/s. The wheel is driven at

24,000 rpm from a 400 Hz excitation source. The gyro wheel is mounted inside a cylinder which serves as the journal of a gas bearing. The cylinder is suspended on the side and ends by a film of gaseous nitrogen emanating from a series of holes in the supporting sleeve. The signal generator, which senses the angular displacement of the output axis, and a torque generator used in initial erection are coupled to the cylinder.

Each pendulous integrating gyro accelerometer (three of which are mounted on the stable table) contains a single-degree-of-freedom gyro. The gyro motor and flywheel are shifted along the spin reference axis to provide the desired pendulosity about the output axis. The gyro is a synchronous hysteresis type similar in construction to the reference gyro but smaller in size. It has an angular momentum of 94,000 g cm^2/s at a wheel speed of 12,000 rpm and is driven by the same 400 Hz source that drives the reference gyros. The accelerometer gyro is also mounted in a gas floated cylinder. The pendulous cylinder is free to rotate about the gyro input axis along which the acceleration is to be measured. The pendulosity causes a torque and therefore a precession proportional to acceleration along the input axis. The speed at which the gyro cylinder rotates is therefore proportional to acceleration and the position is proportional to velocity. An optical incremental encoder on the input axis is used to measure the velocity information.

A significant portion of the platform supporting electronics is required to close the platform gimbal servoloops and the accelerometer servoloops. The servoloops use a 4.8 kHz suppressed carrier modulation system with the signal generator outputs being amplified and demodulated on the gimbals of the platform. The resulting dc signal from the platform is routed to a separate electronics box where it is shaped by a lead-lag stabilization network, remodulated, amplified, and demodulated to drive a dc power bridge which supplies current to the appropriate torquer. Another major function of the supporting electronics is shaping the accelerometer optical encoder outputs. The encoder sine and cosine waves are amplified and converted to square waves for processing in the digital computer system. This system as well as the gimbal readout system, both of which interface very tightly with the digital computer system, is discussed in more detail later.

The supporting subsystems include separate power supplies which derive, from the vehicle 28 V dc buss, all ac and dc voltages necessary to operate the platform system. A three-phase 400 Hz sine wave and three single-phase square wave reference signals at 4.8 kHz, 1.92 kHz, 1.6 kHz, and 56 V dc are provided. Another supporting subsystem is the gaseous nitrogen supply utilized to float the gyro cylinders. Nitrogen is supplied from a 0.056 m^3 (2 ft^3) storage reservoir pressurized to 20.7×10^6 N/m^2 (3000 psi). The gas is regulated to 10.3×10^4 N/m^2 d (15 psid) for use in the platform.

Because of the problems involved in providing redundant stabilizing gyros and other platform elements, the platform does not utilize the extensive redundancy found in some of the other guidance and control subsystems. Instead of providing redundancy at the component or module level, it is more expedient in this case to provide a total system backup. The spacecraft is used to back up the Saturn launch vehicle guidance system during the orbital and translunar injection phases. It will also provide a backup for the Saturn platform as well as the guidance computations performed in the digital computer system. The backup is limited to the later phases since it is not feasible to implement the guidance equations used to inject the vehicle into earth orbit because of limitations of the spacecraft computer memory capacity. Some consideration is being given to a secondary simplified reference system within the launch vehicle (e.g., a strapped-down system) to provide a backup to the platform during all flight phases. Another approach being considered is the provision for manual booster control in the event of a platform system failure. In any case, the launch vehicle digital system must continue to function in all phases regardless of the guidance system backup employed since sequencing, telemetry calibration, and other functions are still performed by the launch vehicle digital computer.

In addition to the total system backup, redundancy is incorporated in certain critical portions of the platform where it can be readily applied. Primary examples of this are as follows:

1. The multispeed analog resolvers on the gimbal pivots, which are used to measure the vehicle angular orientation with respect to the platform.

2. Two channels of information are provided from each optical incremental encoder on the accelerometer. Both the optisyns and signal conditioning circuitry are duplexed. The two channels have equal resolution and provide a redundant channel of information into the data adapter.

3. Duplex redundancy is applied in portions of the circuitry of the power supply package used for excitation of the stabilizing and accelerometer gyros.

Since items 1 and 2 involve very close interfaces with the digital computer system, a detailed functional description of this portion of the guidance and control system, which includes some platform and some digital system elements, is covered here. The accompanying demonstration of reliability improvement through the use of redundancy is also covered on a functional basis rather than as individual elements in separate subsystems. In the overall subsystem reliability assessments, however, the reliability of the individual elements are included in their respective subsystems.

A block diagram of the multispeed resolver channels, including those portions of the digital computer system data adapter used to process the information and provide vehicle attitude correction commands, is shown in Figure 14.

The three resolvers, one for each coordinate axis, have both a 32:1 and a 1:1 winding on the same magnetic structure. For the 32:1 winding, 32 electrical degrees correspond to one mechanical degree. The outputs of the

resolvers are fed through successive platform gimbals by means of sliprings. The resolver excitation frequency (1016 Hz) is derived from the digital computer clock and fed to the platform. Two power supplies are used, and the system is organized such that no 32:1 and 1:1 system in the same channel receives power from the same supply. Therefore, the system is arranged such that a failure in one resolver system or power supply is backed up by the other system.

The outputs of the resolvers are fed to RC phase shift networks in the data adapter and then to crossover detectors (COD) which detect when each signal crosses zero going positive. This signal is then gated to an 11-bit counter in the data adapter. Crossover of one of the sinusoidal signals is used as a start pulse and gates the 2.048 MHz computer clock to the counter. The other sinusoidal crossover is used to stop or turn off the 2.048 MHz counter. Therefore, the value obtained by the counter is directly proportional to the phase shift between the two signals and is representative of resolver shaft position and vehicle attitude. Either a single or a double RC network is employed on the single speed resolver. The 32:1 system employs a double RC network resulting in an equivalent resolution of 64:1. The selection of a single or double RC network for the single speed system is under program control. The single network provides a whole value; however, in case of failure of the 32:1 system, the resolution of the 1:1 system may be doubled (2:1) by employing the additional RC network. (For the 2:1 system to back up the 64:1 system, a decrease in resolution by a factor of 32 must be tolerated.)

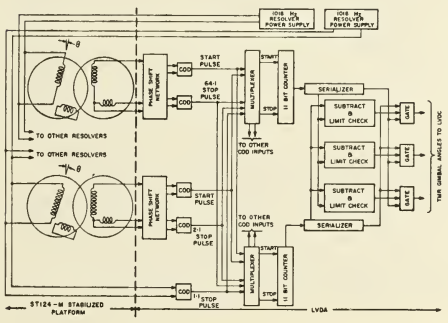


Figure 14. LVDA Gimbal Angle Processing System

The multiplexers in the data adapter are duplexed and all resolver inputs are gated through each multiplexer. The resolver to be read into the duplexed counters is selected by computer program. The output of each counter is routed to three (TMR) subtract and limit check circuits, which compare the counter readings within a predetermined range. The computer is alerted if the subtract and limit test has failed. A counter disagreement indicates either a power supply, COD, resolver, or counter failure. A power supply failure results in multiresolver error readings which may be logically assessed by the computer program. When the subtract and limit test fails to determine if the failure is due to a counter, a pseudo-resolver signal, which is dependent on the computer program, is used to turn on the start and stop signal thereby setting a predetermined value in the counter. If a failure does not occur in this test, it may be assumed that the counters are good and that either a code or resolver error caused the disagreement between the two values. If the error is not corrected within a prescribed period of time or within a given number of iterations, the backup system is employed. If a failure occurs in the counter test, the proper counter and serializer channel may be selected for further use.

For a reliability analysis, this system may be further simplified as shown in Figure 15. Indicated in each block are the functions or hardware grouped together for this analysis. The reliability analysis of the system may be considered in three parts (Fig. 15). The first part uses nonconventional duplexing and consists of the resolver excitation sources, resolvers, platform sliprings, and COD's. The second portion is made up of conventionaly duplexed input multiplexers, counters, and serializers. The third part is the TMR subtract and limit check circuits. The reliability of each part may be considered independently of the others, and the reliability of the system is the product of the reliability of each part.

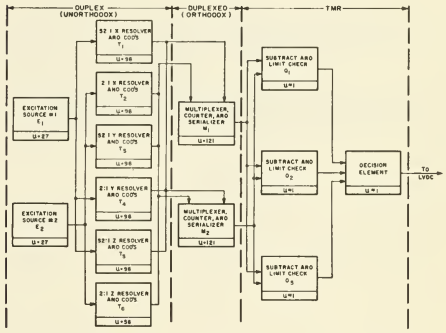


Figure 15. Block Diagram of Gimbal Angle Processing System

For the unorthodox duplex portion to function properly, the following conditions must be met:

$E_1 \cdot \overline{E}_2 (T_1 \cdot T_4 \cdot T_5)$ - one excitation source and three resolvers must be good, or

$\overline{E}_1 \cdot E_2 (T_2 \cdot T_3 \cdot T_6)$ - same as above except the other set of components are considered, or

$E_1 \cdot E_2 \left[(T_1+T_2)(T_3+T_4)(T_5+T_6)\right]$ - both frequency sources and at least one resolver in each axis must be functional.

67-553

When these conditions are treated in detail, the reliability of the unorthodox duplex part is

$$P = R_e^2 [-R_t^6 + 6R_t^5 - 12R_t^4 + 6R_t^3] + 2R_e R_t^3 \qquad (39)$$

where R_e is the reliability of one excitation source, and R_t is the reliability of platform sliprings, the gimbal angle resolver, and the two COD's required for each resolver output.

The reliability of the second and third parts, found by applying equations 10 and 17, is

$$P = 2R_m - R_m^2 \qquad (40)$$

$$P = 3/2\,R_q - R_q^3/2 \qquad (41)$$

where R_m is the combined reliability of one multiplexer, counter, and serializer and R_q is the reliability of the subtract and limit check circuitry. Combining these expressions yields the attitude input system reliability given by

$$P = [R_e^2 (-R_t^6 + 6R_t^5 - 12R_t^4 + 6R_t^3) + 2R_e - R_t^3]$$

$$[2R_m - R_m^2] \left[\frac{3}{2}R_q - \frac{R_q^3}{2}\right] \qquad (42)$$

Generic failure rates for the various components have led to the following subsystem reliabilities for one flight.

$$R_e = 0.999973$$

$$R_t = (R_{resolver})\,(R_{COD})^2 = (0.999914)\,(0.999994)^2$$

$$= (0.999902)$$

$$R_m = 0.999879$$

$$R_q \approx 1.$$

The unreliability of the system can then be calculated to be $U_r = 1$. In comparison, the reliability of a simplex system is given by

$$P = (R_e)\,(R_t)^3\,(R_m)$$

and is found to yield an unreliability of $U_s = 442$. Utilizing redundancy in the system has therefore decreased the unreliability of the system by $U_s/U_r = 442/1 = 442$.

Note that in the system just described an additional decision technique has been used, i.e., the computer logical capability. Previous discussion has been confined to hardware redundancy; however, with this scheme, the computer program and logical capabilities ascertain the system or redundant path to be used for further operation. This type of decision element provides the greatest capability and flexibility; however, complicated programs become even more complex and the normal computational processes are interrupted while this task is performed.

The second portion of the platform employing redundancy, the accelerometer readout channels, also interfaces very closely with the digital subsystem. As was the case with the gimbal resolver channel, these elements are also functionally described and the benefits of redundancy are demonstrated as a single system. A block diagram of the system used in measuring and processing the acceleration information is shown in Figure 16. The figure shows a single measuring channel. Three identical channels are employed to measure the vehicle acceleration along three mutually perpendicular axes.

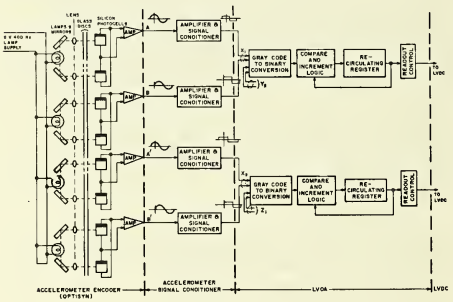


Figure 16. Accelerometer Processing System

The acceleration sensing device is a single-degree-of-freedom gyro unbalanced about its output axis. A torque is produced by the unbalance or pendulosity which is proportional to the acceleration to which the pendulous mass is subjected. The precession angle of the gyro is proportional to the integral of the acceleration. An optical incremental encoder provides a measure of inertial velocity with a resolution of 0.05 m/s^2.

The encoder, which is mounted directly to the gyro head on the platform, contains lamps, mirrors, lens, photocells, and amplifiers. The lamps are excited by a 5 V Hz supply from the platform system. Light from the lamps, which are equally spaced around the periphery of the encoder, is reflected from the mirrors through lenses and passes through two glass discs. Each disc has deposited on it equally spaced opaque lines. Mirrors are used to reduce the number of light bulbs required. One disc is fixed while the others rotate. The light input to each pair of photocells or the input signal to the amplifiers approximates sinusoidal functions as one of the discs rotates relative to the other. The photocells are connected such that maximum signal pickup occurs on one photocell while the other photocell pickup is minimum and vice versa. Effectively, one pair of the photocells

generates the positive portion of the sine wave while the other pair produces the negative portion of the wave. From the amplifier in the encoder on the stable element, the signal is fed through the three platform gimbals by means of sliprings to the accelerometer signal conditioner unit where the signal is further amplified and clipped to obtain square waves.

Two signals (one sine and one cosine) for each channel are fed to the data adapter. These signals represent, in gray code, incremental velocity inputs. One sine wave and one cosine wave are processed in the data adapter logic to give four velocity increments, each increment representing a change in velocity of 0.05 m/s. Before this is used by the computer system, it is converted to a binary number. After the gray code to binary conversion, the 0.05 m/s incremental inputs are summed in a recirculating register in the data adapter. The register is 12 bits plus sign; therefore, a velocity of 204.8 m/s can be accumulated before it overflows. This value is read into the computer approximately once per second, and the entire value velocity of 26 bits stored in computer memory is updated.

Figure 17 indicates the accelerometer readout system organization from a reliability standpoint. The +20 V dc and 5 V 400 Hz power required in the encoder amplifiers and for light bulb excitation is simplex. The voltages to the accelerometer encoders and signals from the platform to the LVDA are fed through platform sliprings. Each block to the immediate right of the power supply consists of the accelerometer encoder (made up of lamps, photocells, an amplifier, and platform sliprings), amplifiers, and signal conditioners located in the platform electronics. The output of each block consists of two signals, a sine and cosine wave, which are necessary to obtain the velocity increments in one axis. (Although magnitude can obviously be obtained from one signal, two signals are necessary to determine direction.) Each part of the gray code to binary conversion is unique to each of these signals and will be lumped with the block on its left for reliability analysis. After conversion to gray code, two accelerations in a different axis are stored together in a register in a glass delay line as indicated in the figure. One of three delay lines can fail without resulting in a system failure; however, other combinations of accelerometer or signal condition failures can result in a system failure. The system is rather complicated to analyze; however, the following general conditions apply:

1. With a failure in either the compare and increment logic, and/or one delay line, one of the other accelerometer signals not associated with the failed logic or delay line can be lost without a system failure; i.e., if in Figure 17 the top channel delay line is lost, X_1 and Y_2 are lost. A failure in either of the Z accelerometer inputs can be tolerated, but a failure in either X_2 or Y_1 results in a system failure. Similar reasoning is appropriate for each of the other channels.

2. With all increment logic and the three delay lines functional, only one accelerometer signal in each of the three axis is required.

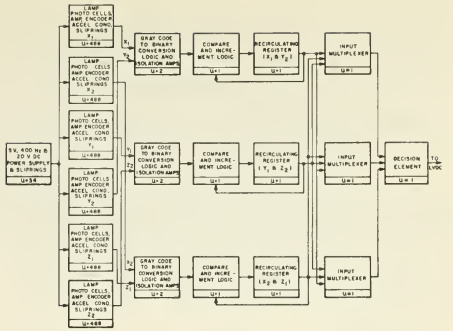


Figure 17. Accelerometer Processing System

The reliability of that part of the system between the power supplies and the input multiplexer is given by

$$P = 3R_i^2 (1-R_i) R_k^2 (2R_k - R_k^2) + R_i^3 (-R_k^6 + 6R_k^5 - 12R_k^4 + 8R_k^3) \quad (43)$$

where R_i is the reliability of one channel of logic including increment logic and the delay line, and R_k is the reliability primarily of the accelerometer encoder and signal conditioning circuitry although it also includes platform sliprings, isolation amplifiers, and gray code to binary conversion logic.

The input multiplexers are conventional TMR and from equation 17 have a reliability given by

$$P = \frac{3R_m}{2} - \frac{R_m^3}{2} \quad (44)$$

where advantage has been taken of failures in opposite directions. R_m is the reliability of a simplex multiplexer.

The reliability of the complete redundant system then is given by

$$P = [R_e] \left[3R_i^2 (1-R_i) R_k^2 (2R_k - R_k^2) + R_i^3 (-R_k^6 + 6R_k^5 - 12R_k^4 + 8R_k^3)\right] \left[\frac{3R_m}{2} - \frac{R_m^3}{2}\right] \quad (45)$$

where R_e is the reliability of the 20 V and 5 V excitation sources and includes those sliprings necessary to get power to the encoders, and all other quantities are as previously defined. Generic failure rates and subsystem analysis yield the following reliabilities for those terms in equation 45.

$R_e = 0.999966$, $R_i = 0.999998$, $R_k = 0.999510$, and $R_m = 0.999999$. Evaluation of equation 45 using these values yields a total system unreliability of $U_r = 35$.

A simplex system would have a reliability given by

$$P = (R_e) R_k^3 R_i^3 R_m. \quad (46)$$

Using these subsystem reliabilities results in a simplex unreliability of $U_s = 1511$.

Comparing the unreliabilities of the redundant and simplex systems indicates a gain factor of 43.2 over the simplex system.

Note that the reliability of the stable elements and the accelerometers was not included in this analysis. Only that part of the system used in processing accelerometer information was included. Since the accelerometers are simplex, an accelerometer failure could result in a system failure.

The value of the computer in recognizing failures is further illustrated in this system. The computer system reads both the prime acceleration and its backup, i.e., X_1 and X_2, etc., and performs a reasonableness test before either is used in the solution of the guidance equation. The computer subtracts the two values stored in the delay lines to determine if the values are consistent or in agreement. If they compare within reasonable limits, either value may be used. If a difference exists, the computer then compares each value with previous values to determine which delta value is more reasonable. The velocity profile of the vehicle can be approximated with a fair degree of accuracy through simulations before flight, and maximum delta velocities expected between successive readings can be determined within reasonable limits.

The third example of redundancy within the platform system is the ac power supply. Although a portion of the circuitry is simplex, duplex redundancy is employed in the oscillator and frequency divider circuitry.

From the primary 28 V dc vehicle power source, the power supply derives the ac power to drive the gyro wheels and provides the excitation voltage for the gimbal synchros and resolvers. A simplified block diagram of the power supply is shown in Figure 18.

The power supply uses a quartz crystal oscillator as a reference. By frequency division, temperature-stable square waves of 19.2 kHz, 4.8 kHz, 1.92 kHz, and 1.6 kHz are derived. The buffered 1.6 kHz and 1.92 kHz square waves are used as reference signals for the platform resolvers. The 4.8 kHz output is routed to the platform electronics assembly where it is utilized in the platform and accelerometer stabilizing circuits. The 4.8 kHz square wave is also used as the reference for a cyclic register, which produces six push-pull 400-Hz square wave outputs in 30-degree increments. The output of this circuitry is transformed and filtered to provide the 3-phase 400 Hz sine wave power which drives the platform gyros.

As indicated in Figure 18, the oscillator, frequency divider, and cyclic register circuits are duplicated. The signal from each channel is fed to the failure detection and switchover circuitry. Both of the duplicated channels are energized, with only one actively controlling the power supply. Any failure in the active oscillator circuitry causing a detectable loss of output voltage will result in an automatic switchover to the standby section.

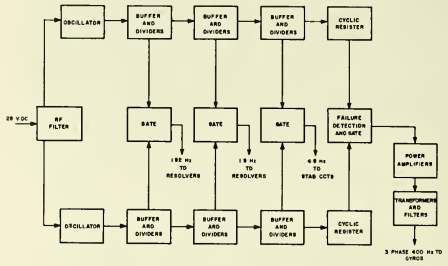


Figure 18. Platform AC Power Supply

To portray the benefits of the redundant circuitry on the overall power supply reliability, a simplified reliability model is shown in Figure 19. The reliability of a single channel of the redundant portion of the system is $R = 0.999845$. Applying equation 9 to the duplex redundant oscillator section yields $P \geq 0.999999$ and $U_{rp} \leq 1$. The equivalent unreliability of the various segments of the power supply is indicated in Figure 19. Considering the duplex oscillator and adding the unreliability of the simplex elements results in $U_r = U_{sp} + U_{rp} = 2 + 356 + 11 + 1 = 370$.

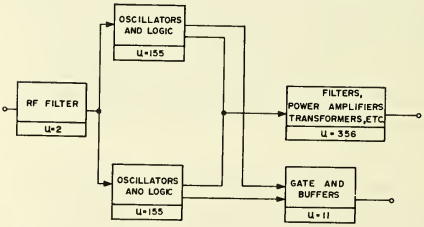


Figure 19. Platform AC Supply (Simplified Model)

If the total power supply including the oscillator was simplex, the unreliability would be

$$U_s = 2 + 155 + 356 + 11 = 524.$$

The overall benefit of redundancy in this case is the reduction of the unreliability by the following ratio:

$$\frac{U_s}{U_r} = \frac{524}{370} = 1.42.$$

The relative improvement is considerably below that obtained in some of the other subsystems because a significant portion of the power supply circuitry could not readily be made redundant.

To demonstrate the overall platform system reliability, the system is assessed by individual elements. The total system consists of six major elements: an inertial platform, a platform electronics assembly, an accelerometer signal conditioner, an ac power supply, a 56 V dc power supply, and a nitrogen gas supply. A block diagram indicating the interconnection of these various elements is shown in Figure 20; the unreliability of these elements for the 6.8 hour mission is indicated. The numbers shown include the reliability improvements in those various elements where redundancy is applied. As shown, the total unreliability of the system including the redundant elements is $U_r = 13,531$. If the system was totally simplex, the following increase in unreliability in the three segments previously discussed would result.

Resolver channels:	$U_s = (3 \times 86)$	$= 258$
Accelerometer readout channels:	$U_s \quad (3 \times 488)$	$= 1464$
Ac power supply:	U_s	$= 155$
	Total increase	$= 1877$

Therefore, the unreliability of a totally simplex platform system would be $U_s = 13,531 + 1877 = 15,408$. The overall system improvement ratio resulting from redundancy is therefore

$$\frac{U_s}{U_r} = \frac{15,408}{13,531} = 1.14.$$

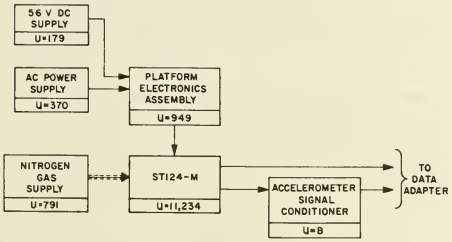


Figure 20. ST124-M Platform System

Figure 21 is a schematic of the guidance system indicating that the launch vehicle platform system is backed up by the spacecraft system. A failure of the Saturn launch vehicle platform is sensed by the digital system by comparing the measured gimbal angle rates with nominally expected values. When an unreasonable signal is read, the digital system operates a light on the astronaut's control panel. In addition, the astronaut has displayed information derived from various spacecraft sensors, as well as communications with ground, from which indication of the system performance can be derived. If a failure or degraded performance of the launch vehicle system is indicated, the astronaut can switch the spacecraft guidance signals directly into the launch vehicle control computer. This implementation does not provide a total backup for the digital system as well as the platform; the digital system must continue to perform many other vehicle functions.

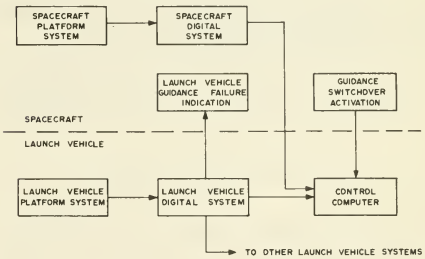


Figure 21. Platform Backup System

To obtain an approximation of the benefit derived from this backup arrangement, it is assumed that the two guidance systems are equally reliable and that the sensing and switching mechanisms are simple and reliable as compared to the overall systems. The unreliability of the launch vehicle platform system can be broken down as (1) through earth orbit injection, $U_{ra} = 8645$, and (2) balance of launch vehicle mission, $U_{rb} = 4886$.

Applying equation 9 to U_{rb} yields

$$(U_{rb})^2 = (4886 \times 10^{-6})^2 \times 10^6 = 24$$

where $(U_{rb})^2$ is the unreliability of the platform systems during the period when the launch vehicle platform is backed up by the spacecraft platform, i.e., from orbital injection to completion of the mission. With guidance backup applied only during the orbital and lunar injection phases, the platform system unreliability consequently is

$$U_r = U_{ra} + (U_{rb})^2 = 8645 + 24 = 8669$$

where U_{ra} is the unreliability of the launch vehicle platform through orbital injection.

Other backup approaches that would further reduce the unreliability have also been mentioned previously. The possibility exists that platform system backup could be provided throughout the launch vehicle flight by a

simplified strapped-down guidance system, or by relying on the astronaut to manually control the system in the event of failure. In the latter case, the astronaut would monitor vehicle angular and translational indications provided in the spacecraft to steer the vehicle into orbit with a degraded accuracy as compared to the primary guidance.

If the same simplifying assumptions as with the spacecraft guidance backup system are made (i.e., $\overline{P}$ (backup) = $\overline{P}$ (primary), and $\overline{P}$ (sensing and switching) = 0), the following unreliability results from equation 9.

$$U_r = (U_{rab})^2 = (13,531 \times 10^{-6})^2 \times 10^6 = 183$$

where $(U_{rab})^2$ is the unreliability of the system when the launch vehicle platform is totally backed by another system during all phases of launch vehicle operation.

A resume of the platform system unreliability and the benefits of the backup schemes are shown in Table IV,

Table IV. Platform System

	U_r	U_s	U_s/U_r
Launch vehicle system	13,531	15,408	1.14
With backup out of orbit	8,669	15,408	1.77
With proposed total flight backup	183	15,408	84.30

which indicates that a very significant reduction in unreliability in the platform system can be obtained only by providing a backup throughout the total flight. Therefore, several total backup approaches are being pursued.

Control System

For a logical functional description and practical redundancy application, the control system is broken down as attitude rate sensing, multiengine (S-IC and S-II) stage propelled phase control, single-engine (S-IVB) propelled phase control, and S-IVB coast phase control.

The rate sensing system is composed of two boxes containing the rate sensors (the rate gyro package) and the associated electronics (the control signal processor package). The rate gyro package contains nine rate gyros so arranged that angular rate about each of the vehicle axes (pitch, yaw, and roll) is sensed by three separate instruments; thus three separate signals, independently derived, are available for each axis. The individual rate gyros are single-degree-of-freedom instruments containing a spin motor which operates at a synchronous speed of 24,000 rpm and has an angular momentum of 30,000 g cm^2/s. Angular rates about the input axis, which is aligned with the vehicle pitch, yaw, or roll axis depending on the case mounting direction, are sensed by a 400 Hz microsyn pickoff that is electromagnetically coupled to the gyro gimbal. The output of the microsyn is proportional to the vehicle angular rate about the input axis. The microsyn outputs, one from each of the three instruments in each axis, are fed in parallel into the control signal processor. Nine demodulator modules, three for each axis, receive the rate gyro error signals. Each demodulator module amplifies the input signal and provides a plus or minus dc voltage proportional to the ac input amplitude.

The power for the rate sensing system is obtained from three separate 28 V battery supplies over three busses. Three static inverters in the control signal processor supply the 26 Vrms 400 Hz power to the rate gyros and demodulators; likewise, three dc power circuits supply the necessary 60 volts to the demodulators. Each primary power buss with its associated inverter and dc power circuit supplies three gyros and associated electronics; one in each of the pitch, yaw, and roll groups.

The rate sensing utilizes the PRS form of redundancy. A simplified diagram of one channel neglecting power supplies is shown in Figure 22. The rate signal outputs from the primary command demodulator and the reference demodulator are sent to a comparator, which consists of two differential amplifiers, an amplitude sensor, and a relay driver. If the difference between the primary and reference channels exceeds a preset level, the comparator circuit operates relays which switch the primary channel out of operation and substitute the standby channel into the primary command position. Thus, if a malfunction occurs in either the primary or reference channels, the standby channel will be substituted. If a malfunction occurs in the standby channel with the other channels performing properly, no switching occurs. The difference level, at which the circuit switches (1.65 deg/s), is determined from compromise considerations of hardware tolerance characteristics and expected vehicle motions. The reference channel serves solely as a reference and is never used to provide the rate command to the remainder of the system. The PRS redundancy as implemented in this subsystem does not provide the capability of switching back during flight to the primary channel after the standby channel has been substituted. Such an arrangement causes the subsystem to revert to an equivalent simplex system after a single discrepancy, even if it is transient in nature. A multiple switching capability such as that utilized in the digital system would be more reliable, but would also be more difficult to implement in an analog system.

The reliability assessment of the individual blocks shown in Figure 22 is

$$R = (R_{demod})\ (R_{rate\ gyro}) = (0.999919)\ (0.998453)$$

$$= 0.998372$$

$$R_v = 0.999931.$$

These numbers are applicable to the total flight time, since the rate system must function throughout flight. Applying these numbers in equation 18 for the PRS system yields P = 0.999994; or, expressed in terms of malfunctions per million flights, $U_r = 6$. The numbers shown in

Figure 22 express the unreliability of the individual channels of the PRS system. For the equivalent simplex system, $U_s = 1628$.

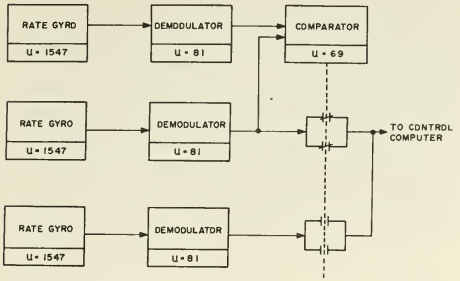


Figure 22. Attitude Rate System (Single Channel)

If the rate control system was composed of three uncoupled control axes, the unreliability of the total system could obviously be obtained by multiplying the above redundant unreliability (U_{PRS}) by three. The three axes are independent except for the internal power supplies. If the internal power supplies are considered, however, the treatment is not quite so straightforward since each of the three power supplies drives one channel in each of the pitch, yaw, and roll axes. A simplified block diagram of the complete three-axis system with the power supply interconnection arrangement is shown in Figure 23. The expression applicable to the total three-axis system shown is

$$R = R_e^3 [(R^3 - R^2)(1 - 2R_v) + R] + 3(1-R_e) R_e^2 R R_v [(1-R_v) - R(1-2R_v)] \quad (47)$$

where R_e = reliability of one of the three power supplies (inverter and 60 V dc supply, combined) and the other terms are as previously defined. The first term in the expression represents the probability of all outputs being good when all three power supplies are assumed to be good. The second term represents the combination of properly functioning situations which result when the power supplies are assumed to fail singly. When two or more power supplies are lost simultaneously, a failure results in either pitch, yaw, or roll. In equation 47,

$$R_e = (R_{\text{inverter}}) (R_{\text{dc supply}}) = (0.999897)(0.999915)$$

$$= 0.999812$$

and R and R_v are as previously indicated.

For the total subsystem, $P = 0.999980$ and $U_r = 20$. If the system was simplex, $U_s = 4972$ would apply. Therefore, the improvement through redundancy is $U_s/U_r = 248$.

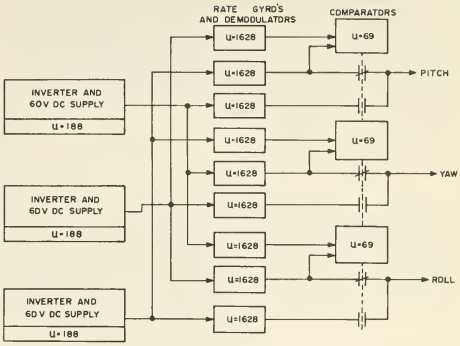


Figure 23. Attitude Rate System (All Channels)

The vehicle attitude rate signals derived in the subsystem, as well as the desired vehicle attitude derived from the stabilized platform and digital computer, are utilized to direct and stabilize the vehicle. The flight control computer processes these input signals and derives in an analog manner the appropriate command signals for the gimbaled engine actuators and auxiliary thruster valves to torque the vehicle as required. Control torques on the first two stages (S-IC and S-II) are derived by positioning the four gimbaled engines on each stage. The control torques for the upper stage (S-IVB) are obtained by gimbaling the single main engine and activating the six fixed-direction auxiliary engines. The two techniques are different in basic layout and are discussed separately.

A layout of the control system of the multiengine stages is shown in Figure 24. There are six inputs to the control computer, an attitude and attitude rate for each of the three axes. These signals are individually scaled, filtered, and then routed to the appropriate servoamplifiers which drive the engine actuators. The elements of particular interest in this chain are the filters, or shaping networks, the servoamplifiers, and the servoactuators. The characteristics of each shaping network are those required to satisfactorily provide the required stability margins, taking into account the vehicle structural bending, propellant sloshing, and transfer functions of the remainder of the guidance and control system. In this module, compensation is made for variation between individual vehicles and individual missions. Extensive analysis is required to derive the shaping networks for each particular mission. This particular module, along with its associated isolation amplifiers, is simplex in each of the two multiengine stages. The simplex approach was chosen in this case for two reasons. First, since the mission time of each

multiengine stage is relatively short in duration and the shaping networks are composed of only a few components, the predicted reliability is high even for the simplex version. The second reason is one of engineering compromise to conserve weight and space. Although simple in configuration, the networks are bulky compared to other modules of the control electronics because of the large size of some of the electronic components (capacitors and inductors) required to accomplish the necessary shaping at the low control bending mode frequencies in the range of 0.5 to 5 Hz.

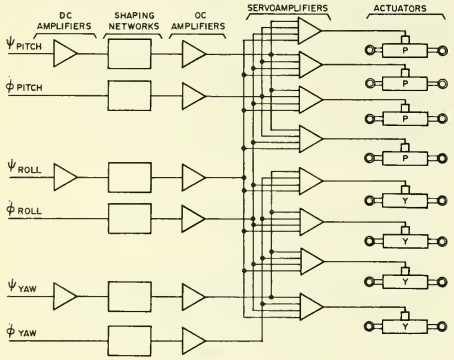


Figure 24. S-IC or S-II Stage Control System

Each servoamplifier is composed of a magnetic mixing amplifier, followed by transistor stages which provide the necessary power gain. A number of inputs are in each magnetic amplifier which is the point in the system where the attitude and rate signals from the appropriate axes are combined. The various input signals into the magnetic amplifier are galvanically isolated from each other since each is applied to a separate winding. The excitation signal for the magnetic amplifier is derived from a chopper-stabilized inverter which converts the dc source to a one kHz signal. The outputs of the eight servoamplifiers drive eight corresponding servoactuators which position the four gimbaled engines as required. These hydraulic servoactuators and the associated fluid supplies make up the other major elements in the multiengine control system. The hydraulic systems of the S-IC and S-II are designed differently to satisfy the individual stage requirements.

A simplified schematic of the S-IC hydraulic servoactuator is shown in Figure 25. The servoactuator receives from the servoamplifier an electrical signal which represents the desired engine position. The electrical signal is applied to the servovalve torque motor, causing a pressure differential to exist between two orifices. This pressure differential positions a spool which in turn regulates the flow in a manner to control the position of a second spool. The flow regulated by this second spool determines the actuator piston location and, therefore, the gimbal angle of the attached engine. The entire system is essentially a three-stage hydraulic power amplifier. In addition to providing the necessary power amplifications and conversion, the servoactuator must also meet certain dynamic response, load damping, and stiffness requirements. These features are provided by hydraulic pressure feedback and shaping within the actuator. The servoactuator also employs the principle of mechanical feedback, which improves reliability by eliminating the need for actuator position information to be electrically sensed and fed to the control computer over long lines through multiple interfaces. The feedback mechanism converts the rectilinear motion of the actuator to a force which counteracts the electromotive force of the input signal on the first stage of the servovalve.

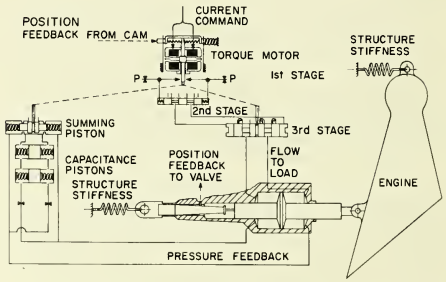


Figure 25. Schematic of S-IC Servoactuator

The actuator has a stall load of 507,000 N (114,000 lb) which is equivalent to a torque of 810,000 Nm (600,000 ft lb) as applied to the gimbaled engine. The expected operating torque range is 540,000 Nm and below; the major torques to be overcome are contributed by propellant duct loads and a thrust vector which does not pass through the center of the engine gimbal bearing. In the presence of these and other loads, the actuator can position the engine through an angle of ±5.2 degrees at a rate of 5 deg/s.

The fluid supply for the S-IC servoactuator is RP-1 fuel taken directly from the turbopump which also supplies the main engine. This makes an extremely simple and reliable onboard hydraulic supply since only filters and interconnecting ducting must be added to the propulsion distribution system. The individual gimbal systems are independent because the turbopump on each engine furnishes the supply for the actuators on that engine.

The S-II servoactuator is functionally similar to the S-IC servoactuator although physically much smaller. The hydraulic flow rates required to position the engine are much lower, so only two stages of hydraulic amplification are required. Mechanical feedback, pressure feedback, and hydraulic shaping are also employed in this actuator. The S-II actuator has a stall load of 202,000 N (45,500 lb). The maximum load expected

to occur during flight is 133,000 N (30,000 lb). The S-II (J-2) engine is gimbaled through an angle of ±7.3 degrees at a rate of 10 deg/s.

The S-II fluid supply is different from the S-IC in that a closed high-pressure system is utilized. The hydraulic power source is a pump driven by the turbopump shaft on each gimbaled engine. The other major components in the fluid supply are an accumulator, which supplies flow to supplement the main pump during periods of peak demands, and a low flow auxiliary pump.

To illustrate the reliability improvement afforded by the multielement control on the first two stages, a block diagram of one-axis control neglecting the hydraulic supply is shown in Figure 26. The case illustrated is S-IC or S-II pitch control; the shaping networks and associated amplifiers are simplex. The servoamplifiers and actuators are representative of the inherent MPE redundancy.

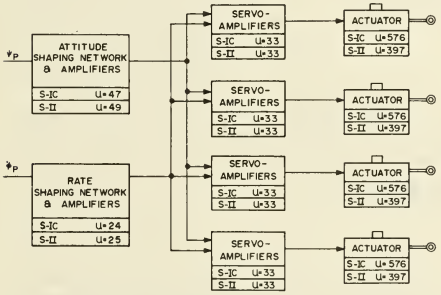


Figure 26. S-IC or S-II Stage Pitch Control System

If one element of the MPE configuration can fail without a loss of the mission, the unreliability is dramatically reduced compared to a system requiring all elements to function. This capability can be designed into a system by a certain overdesign as compared to a nominal failure-free situation. For instance in the gimbal system under discussion, an additional gimbal angle and gimbal rate capability must be provided. Structural and aerodynamic aspects must also be considered. During certain times of flight and under certain combinations of adverse conditions, the S-IC and S-II stages cannot be controlled with a failure in one element of the MPE configuration. In a precise analysis, the probability of loss of mission in the event of a failed channel during the various flight phases would have to be considered. The capability of maintaining control when a channel is lost exists during an appreciable portion of the flight; however, the simplifying assumption is made here that MPE redundancy exists throughout. With this assumption, the reliability of the MPE portion of the subsystem for the S-IC stage can be found from equation 21, where

$$R = (R_{electronics})\ (R_{actuator}) = (0.999967)\ (0.999424)$$

$$= 0.999391$$

resulting in $P = 0.999997$ and $U_{rp} = 3$. For the simplex portion of the electronics,

$$R = (R_{\Psi})\ (R_{\varphi}) = (0.999953)\ (0.999976)$$

$$= 0.999929$$

and $U_{sp} = 71$. Adding the simplex and redundant portions yields $U_r = U_{sp} + U_{rp} = 71 + 3 = 74$.

The probability of failure for the simplex elements and individual parallel elements is shown in Figure 26. If all elements must function properly (i.e., if no inherent redundancy exists), the unreliability is found by adding the unreliability of all elements; thus $U_s = 47 + 24 + 4(33 + 576) = 2507$.

The layout for the yaw channel control is similar to that for pitch. Except for a slight difference in the shaping networks, the circuits for the two channels are identical. As shown in Figure 27, the roll signal is mixed with the pitch and yaw signals in all eight channels. The applicable unreliability numbers are also shown. In addition to the numbers developed, Table V shows the unreliability of the complete S-IC pitch, yaw, and roll control system (electronics plus actuators) for the implemented redundant system as well as a corresponding simplex system.

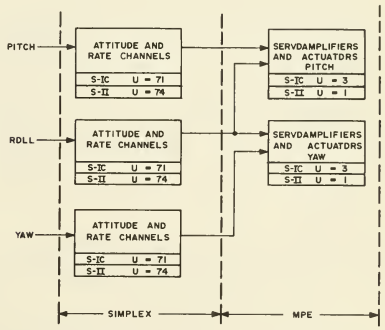


Figure 27. S-IC or S-II Stage Control System (Simplified Model)

The hydraulic fluid supply systems have not been included in this assessment. The layout of the S-IC and S-II hydraulic systems is such that the fluid supply attached to each engine drives the pitch and yaw actuator of that engine. A block diagram of the overall stage gimbal system and the unreliability of the individual blocks is shown in Figure 28. If a loss of one MPE

channel in both pitch and yaw can occur simultaneously without loss of control, which is consistent with the assumptions previously made, the capability of loss of one fluid supply out of the four also exists. Since the general equations cannot be applied directly to a multi-element system having this interconnection arrangement, a specific equation has been derived for this multi-element layout. With the assumptions stated, the following expression results:

$$P = R_j^8 R_w^4 + 8R_j^7 (1-R_j) R_w^4$$

$$+ 4R_w^3 (1-R_w) R_j^8 + 16(1-R_j)^2 R_j^6 R_w^4 \qquad (48)$$

$$+ 8(1-R_j) R_j^7 (1-R_w) R_w^3$$

where R_j = individual actuator-servoamplifier reliability as previously indicated and R_w = individual hydraulic supply reliability = 0.998581 for the S-IC stage. Inserting the reliability numbers into equation 48 and reconverting results in the total subsystem assessment shown in Table V. The overall improvement ratio resulting from redundancy is $U_s/U_r = 43.1$.

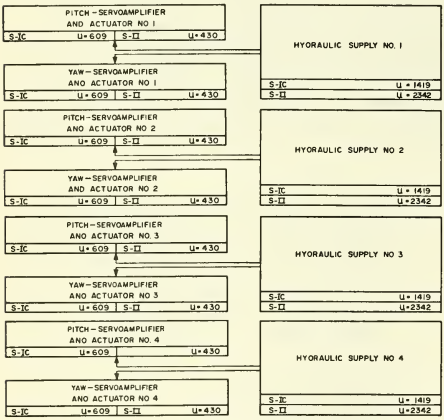


Figure 28. Multiengine Stage Gimbal System

Table V. S-IC Control

	U_{rp}	U_{sp}	$U_r + U_n$	U_s	U_s/U_r
Pitch, neglecting fluid supply	3	71	74	2507	33.9
Pitch, yaw, and roll, neglecting fluid supply	6	213	219	5085	23.2
Pitch, yaw, and roll, including fluid supply	37	213	250	10,760	43.1

Note that this analysis does not specifically consider the effect of "engine-out," i.e., the loss of propulsion of one of the four control engines. Even though the direct effect on the control system which would be the loss of control torques derived from one actuator in each axis is considered, other interactions are not treated in this simplified analysis.

As previously mentioned, the basic layout of the S-II control system is similar to that of the S-IC. Except for shaping networks, the electronics for the S-IC and S-II are identical, with the outputs of the servoamplifiers being switched at staging. Figures 26, 27, and 28 apply also to the S-II stage and show the corresponding unreliability numbers for the individual major elements for both stages. Similarly, equations 21 and 48 are used in the reliability assessment. The numbers used in the S-II stage assessment are $R_j = 0.999570$ and $R_w = 0.997658$. The results are shown in Table VI. The overall improvement ratio in the S-II stage through redundancy is $U_s/U_r = 46.8$.

Table VI. S-II Control

	U_{rp}	U_{sp}	U_r	U_s	U_s/U_r
Pitch, neglecting fluid supply	≤1	74	75	1794	23.9
Pitch, yaw, and roll, neglecting fluid supply	≤2	222	224	3662	16.3
Pitch, yaw, and roll, including fluid supply	61	222	283	13,254	46.8

The layout of the propelled phase pitch and yaw control of the S-IVB is basically different from that of the multiengine stages. Since only one main propulsion engine is employed, the control torques are derived by positioning a single actuator in each axis. A layout of the pitch and yaw control system is shown in Figure 29; the layout of the pitch and yaw channels is similar. Control about the roll axis is maintained by auxiliary engines and is discussed later. The electronic modules in the pitch and yaw channels are similar to those previously discussed, with the identical modules employed in the first two stages being used where possible. The shaping networks are different since they must have the particular characteristics required to stabilize the S-IVB stage. The S-IVB servoactuator is very similar in design to that previously described for the S-II stage, although a few features differ to adapt to the particular stage requirements. The S-IVB hydraulic fluid supply is also similar in layout to that of the S-II, but the individual components are of a different design. The major components of the inflight fluid supply system are the engine pump, a motor-driven auxiliary pump, an

integrated accumulator reservoir module, and associated interconnecting tubing.

Because the S-IVB pitch and yaw control torques are derived from a single engine, redundancy is employed to the maximum extent feasible. As shown in Figure 30, PRS redundancy is employed to derive the control signals to the servoactuators.

The reliability numbers applicable to a single PRS channel and the comparator electronics are $R = 0.999657$ and $R_V = 0.999932$. The corresponding

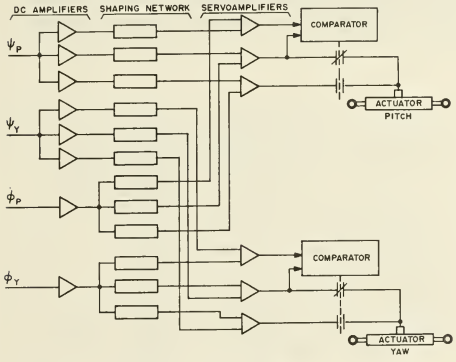


Figure 29. S-IVB Propelled Phase Pitch-Yaw System

unreliability numbers are shown in Figure 30. From equation 18, $P \geq 0.999999$ and $U_r \leq 1$. The reliability numbers for the simplex portions of the system are $R_{actuator} = 0.998437$ and $R_{supply} = 0.993563$. The equivalent unreliability numbers for these elements of the system are also shown in Figure 30. A single hydraulic fluid supply drives both the pitch and yaw actuators. Figure 31, a simplified block diagram of the total system, shows the unreliability associated with the various portions of the system, including the PRS redundant electronics. The resulting composite numbers are shown in Table VII.

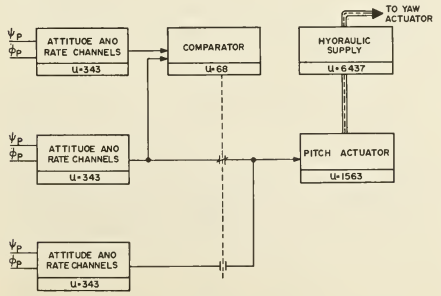


Figure 30. S-IVB Propelled Phase Pitch System (Simplified Model)

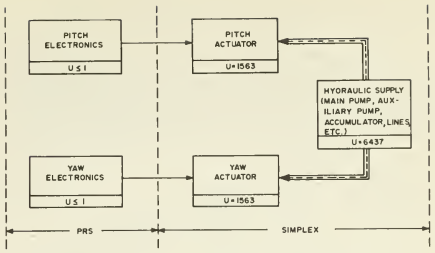


Figure 31. S-IVB Propelled Phase Pitch and Yaw Gimbal System Layout

Table VII. S-IVB Propelled Phase

	U_{rp}	U_{sp}	U_r	U_s	U_s/U_r
Pitch, neglecting fluid supply	≤1	1563	1564	1906	1.22
Yaw, neglecting fluid supply	≤1	1563	1564	1906	1.22
Hydraulic supply	0	6437	-	6437	-
Total system	≤2	9563	9565	10,249	1.07

Thus, for the total system, the unreliability has been decreased by the following factor through redundancy:

$$U_s/U_r = \frac{10,249}{9,565} = 1.07.$$

The numbers reveal a relatively small gain obtained by the redundancy applied in this subsystem; however, redundancy was applied only to the electronics, which is already the most reliable portion of the subsystem. This design is the result of engineering compromise. PRS redundancy was easily applied in the electronics; significant portions already existed in the control computer because of the multiengine stage requirements. On the other hand, the servoactuator and the hydraulic supply were not made redundant because of complexity of implementation and the resulting weight penalty.

This subsystem has a high unreliability because of the major simplex items and the possibility of introducing more redundancy is being pursued. A certain redundancy not considered in this analysis exists in the fluid supply because the system has two pumps. Although the auxiliary pump has a much lower flow than the main pump, it might sustain the system under certain main pump failure conditions. The addition of a second higher flow pump is being considered.

Also being considered is the use of a modified actuator design, which incorporates a "majority-voting" servovalve and essentially consists of a triplication of the valve and mechanical feedback mechanism in the servoactuator. In case of a malfunction in one channel, the two correctly operating channels overpower the third and the system continues to function properly. A considerable improvement could be expected in the valve

and feedback portions of the actuator; both contribute significantly to the actuator unreliability.

As previously mentioned, control about the S-IVB roll axis during propelled flight and about all axes during the coast phase is maintained by torques derived from the on-off operation of six auxiliary thrusters. A layout of the auxiliary control system is shown in Figure 32.

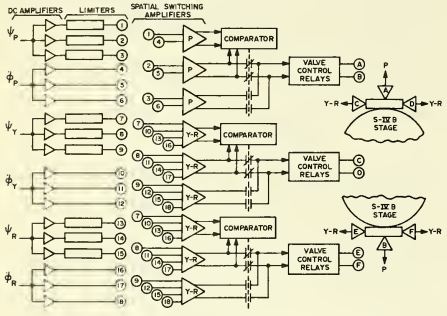


Figure 32. S-IVB Auxiliary Control System

The six inputs to the system (attitude and attitude rates) are derived in the same manner as during propelled phase control. The outputs of the electronic system actuate relays which operate the valves of the six auxiliary thrusters. As indicated, pitch is controlled by engines A and B; yaw and roll signals are intermixed and determine the operation of engines C, D, E, and F to maintain control about these two axes. In addition to scaling amplifiers similar to those employed in other flight phases to establish the correct relative gains in the system, the electronics also include attitude signal limiters, spatial switching amplifiers which operate the propellant valve relays, and spatial comparators. The attitude and attitude rate signals are summed in a magnetic amplifier stage similar to that employed in the propelled phases. The switching function is accomplished in a Schmitt trigger circuit which furnishes the input to the relay drivers. The relay drivers operate double-pole double-throw relays which switch power to the coils of the fuel and oxidizer valves of the thrusters. Pseudo-rate modulation circuitry, which provides a refinement of the simple on-off spatial attitude control techniques, is also included in the spatial amplifier module. The pseudo-rate circuitry provides a modulated band in which the duration and frequency of thruster pulses are varied depending on the input signals. When the input signal exceeds a certain level, the thrusters are commanded to the on position continuously; below a certain level, the thrusters are turned off and the vehicle attitude coasts within the prescribed deadband. The pseudo-rate modulated band provides a more rapid damping-out of disturbances and hence a more efficient utilization of thruster propellants. The electronics also contain a circuit which insures that when a thruster is activated it stays on for a certain minimum time. This characteristic is necessary to maintain the thruster specific impulse at the desired level.

The auxiliary control system employs two types of redundancy: (1) PRS redundancy similar to that previously described is employed in the electronics portion of the system, and (2) the propellant valves of the thrusters are connected in a quadruplex arrangement and are activated by parallel relays.

A simplified diagram of one axis of the auxiliary control system is shown in Figure 33. The pitch axis coast control represents the simplest layout. Roll and yaw coast control are similar to pitch with the exception that they are coupled and require four thrusters. The propelled phase roll control layout is similar to that of Figure 33 except that four thrusters are involved, with two being simultaneously activated for each roll correction.

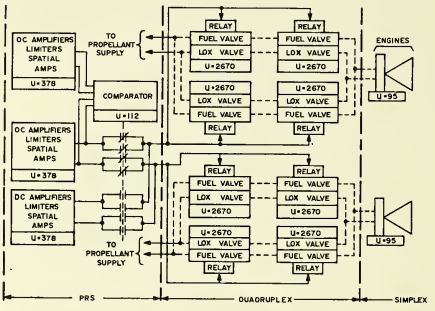


Figure 33. S-IVB Pitch Coast Control System

The reliability numbers applicable to the PRS electronics modules are R = 0.999622 and R_v = 0.999888. From equation 18, P ≥ 0.999999 and $U_{r(elect)} \leq 1$.

The numbers used for a single module of the quadruplex valving arrangement are $R = (R_{relay})(R_{valves}) = (0.999990)(0.997330) = 0.997340$. Since the relay reliability is very high compared to the valves, the simplifying assumption is made that the relay can be included with the valves in this analysis. Using the numbers in equation 20 for a quadruplex arrangement yields R = 0.999986, $U_{r(valves)} = 14$. The total unreliability of the redundant portion is $U_{rp} = U_{r(elect)} + 2\,U_{r(valves)} = 1 + 2(14) = 29$. The term $2\,U_{r(valves)}$ arises from the two sets of quadruplex valves. Except for the valves, the APS engine is simplex. The applicable reliability number is R = 0.999905. Since two simplex engines are employed, the unreliability of this portion is $U_{sp} = 2(95) = 190$. The corresponding unreliability numbers for the individual electronic modules, valves, and engines are indicated in Figure 33. The unreliability of the pitch system as indicated is $U_r = U_{rp} + U_{sp} = 29 + 190 = 219$.

The corresponding numbers for the yaw and roll channels can be similarly derived. These results along with those for the total system, an equivalent simplex system, and the improvement ratio are shown in Table VIII. The total system improvement ratio is $U_s/U_r = 27$.

Table VIII. S-IVB Auxiliary Attitude Control

	U_{rp}	U_{sp}	U_r	U_s	U_s/U_r
Pitch	29	190	219	5908	27.0
Yaw and roll	58	380	438	11,816	27.0
Total system	87	570	657	17,724	27.0

Note that the preceding assessment does not include the simplex APS propellant supply modules, which also supply the propellant for the S-IVB ullage engines.

The unreliability assessment for the various subsystems and the total control system is summarized in Table IX. As previously indicated, the improvement ratio is very large in the case of the rate system which is totally redundant; however, it is not very significant in the S-IVB propelled phase gimbal system because of the simplex hydraulic system. Primarily because of the relatively high unreliability of the latter, the improvement ratio of the total control system through redundancy is only a modest 5.3.

Table IX. Control System

	U_{rp}	U_{sp}	U_r	U_s	U_s/U_r
Pitch, yaw, and roll attitude rate sensing subsystem	20	0	20	4972	248
S-IC pitch, yaw, and roll subsystem	37	213	250	10,760	43.1
S-II pitch, yaw, and roll subsystem	61	222	283	13,254	46.8
S-IVB pitch and yaw propelled phase subsystem	≤2	9563	9565	10,249	1.07
S-IVB auxiliary attitude control subsystem	87	570	657	17,724	27
Total control system	207	10,568	10,775	56,959	5.3

Problems Associated With Redundant Applications

The benefits to be derived from redundancy have been demonstrated, and it has been shown that the unreliability of a simplex system can, in some cases, be reduced by orders of magnitude when redundancy is applied. Although the disadvantages of redundancy are not readily assessed quantitatively, it is recognized that this gain in reliability is at the expense of other design factors or operational procedures. Some of the problems encountered with the application of redundancy are enumerated and the effect of redundancy on the system is indicated.

The most significant disadvantage of redundancy is the increased complexity, both in terms of component parts and system organization. In the simplest forms, i.e., series or parallel components, the number of components is twice that of a simplex system. In a duplex modular system, the number of components is more than doubled to provide a means of sensing and switching. In the triple-modular redundant digital system, the voters and failure isolation and detection circuits require almost as many component parts as a single channel; therefore, a system contains between three and four times as many parts as a simplex system. In the PRS circuitry of the control system, the component count also ranges from three to four times that of a simplex channel, depending on the relative complexity of the comparator required. In the example of the quadruplex redundancy cited in the auxiliary control valves, no sensing or voting was required so the system is four times as complex. The only application of redundancy which does not add additional components to the system is the MPE gimbal system; the complexity of this system was imposed by other design considerations and the benefits of redundancy are achieved without additional complexity. The gain in reliability through redundancy is, in this case, a bonus rather than the primary purpose of the multiple parallel elements.

Other major problems inherent in redundant applications are failure detection and isolation. Failures in the redundant element must be detected and removed before flight. Failure to verify that all channels are operating can actually result in a degradation of the system compared to a simplex system. For instance, consider one trio of a TMR system. If the vehicle is launched with one of the units out, the system would fail if either of the other two malfunctioned. Since there are two remaining units, either of which could result in a system failure, the unreliability of the system is nearly twice that of the simplex system. Failures occurring during flight must be detected so that corrective action can be taken for future flights. Because more component parts are employed in redundant systems, the number of component part failures can be expected to be greater than those in a simplex system by a factor of the ratio of the number of component parts in a redundant system to the number of component parts in a simplex system.

The LVDC, LVDA, and control system illustrate how the failure detection and isolation problem is approached in the Saturn V guidance and control system. In the LVDC and LVDA, disagreement detectors indicate when a failure has occurred in either of these units, each of which consists of over 100 detectors. Several detectors are "OR'ed" together and fed to a bit in a 26-bit register, storing failure indications which can be read by the ground launch computer before flight and telemetered during flight. Sixteen bits of the register are used to store LVDC failure information while the remaining 10 bits are used for the LVDA. Because of the "OR'ing" operation, however, it is not always possible to pinpoint the cause of failures.

To assist in failure isolation before flight, means to switch in and out various redundant paths must be provided. For example, in checking the LVDA power supplies, switching both the feedback amplifiers and the converters is necessary. In the TMR logic, module as well as channel switching is desirable such that a failure can be isolated to two or three logic pages. These

features have been incorporated in the computer system and means are available for checking all alternate paths. The presence of the multiple channels within the redundant system, along with the isolation capabilities incorporated, considerably enhances the troubleshooting possibilities. This is a significant by-product of redundancy, particularly in a complex system such as the Saturn digital system.

Signals from the ground can be substituted during checkout for each of the three inputs to the PRS systems employed in the control system. The comparator's ability to switch can consequently be checked for various combinations of inputs, and the standby units can be exercised. Means are available to switch back from the standby to the prime unit from ground control in case the redundant circuit switches because of an intermittent condition during prelaunch checkout. In addition, the state of the comparators is telemetered such that switching to the standby is detectable during flight; however, the switch-back capability is not present during flight.

The necessity for failure detection, isolation, and removal of failed units is perhaps obvious; however, a more subtle problem arises in using these schemes in an operational prelaunch checkout system. For example, if a failure in a redundant flight item occurs hours before the flight, a spare may be substituted without impact on the countdown or launch. However, should a failure in a redundant item occur just seconds before the scheduled liftoff, the removal of the failed item would require a hold or a scrub, possibly resulting in a costly schedule delay. A tradeoff must be made between the effect of the failed unit upon mission success and the cost, schedule, and other critical considerations brought about by a hold or scrub. It is imperative then that redundancy considerations be included in launch ground rules, where practical. When applied, such considerations complicate launch procedures; when not applied, considerable pressure is brought to bear on engineering judgement.

To derive maximum benefit from the redundancy employed in the Saturn vehicle system, the computer system is utilized to the greatest extent possible because it is the only item within the vehicle capable of making logical choices and decisions. The Saturn V flight program is designed to make maximum use of the existing redundancies in the vehicle hardware. It is generally accepted that a major effort in any guidance and control system is the preparation and checkout of the flight program. This is particularly true in space vehicles where each mission is different from the previous one. Consequently, the "canned" programs cannot be used. Adding redundant features to system hardware complicates flight programs since backup paths or redundant loops must be incorporated. Examples have previously been cited of the value of the computer system in determining "reasonable" values for accelerometer and gimbal angle readings. If it is determined that these values are not "reasonable," alternate modes of operation are followed. Therefore, means must be provided in the various program checkout facilities where failures can be induced and alternate program modes can be checked in a manner similar to that employed in hardware checkout. A problem also exists in determining "reasonable" values, both in terms of which quantities should be used as well as the limits applied to each quantity.

Other disadvantages of redundancy, which are a direct outgrowth of increased complexity, are the physical quantities of increased power, weight, and cost. These quantities have not been, and probably cannot be, accurately assessed, but estimates can be made. The most straightforward of the above quantities to consider is power, since it is reasonable to assume that the power requirements of a system are directly proportional to the number of component parts; i.e., the ratio of the power required by a redundant system as compared to a simplex system may be estimated to be directly proportional to the ratio of the number of components in the two systems.

The weight penalty of a redundant system is not as easy to estimate, for consideration must be given to packaging density and efficiency, heat dissipation, and type of packaging technique employed. In general, weight ratio is estimated to be less than the component part ratio. How much less depends on factors such as type of redundancy employed, failure detection and isolation schemes, packaging techniques, and type of cooling method utilized.

The impact of redundancy upon cost is most difficult to analyze for it runs the gamut of the aforementioned problems. Cost is influenced by the number of parts, system design, checkout, programing, and launch costs. Cost is also greatly dependent upon the type of redundancy employed. From the initial design phase through the launch phase, the cost of a redundant system is probably from 3 to 10 times that of an equivalent simplex system.

Conclusions

The various types of redundancy employed in the Saturn guidance and control system and typical numbers demonstrating the improvements gained have been presented. Although the various types of redundancy show a theoretical difference in the relative improvements, the choice of the type employed in each case is in actuality dependent on the practical implementation aspects. In the design of the Saturn guidance and control system, the following approach was employed: Those portions of the system to which redundancy could be readily applied were first identified and then the type of redundancy was selected by numerical analysis and engineering tradeoff with emphasis on the latter.

Table X summarizes the unreliabilities of the guidance and control system, consisting of the three major systems.

Table X. Summary of Guidance and Control Subsystem Reliability

	U_r	U_s	U_s/U_r
Stabilized platform system*	8,669	15,408	1.77
Digital computer system	282	12,498	44.3
Control system	10,775	56,959	5.3
Total	19,726	84,865	4.3

*U_r includes considerations of backup out of orbit, while U_s refers to a totally simplex system (no subsystem redundancy and no backup).

The digital system, which is almost totally redundant, has a significantly lower unreliability than the other two systems. This should not be interpreted to mean, however, that incorporating redundancy in the other systems is to no avail. Actually, significant improvements are made in all three systems through redundancy. The fact that the three major systems have significantly different reliability and that the most reliable system (in simplex form) employs the highest degree of redundancy emphasizes the philosophy employed in the design of the Saturn system. The approach did not attempt to enforce equal reliability for subsystems of similar significance and complexity; it was instead to benefit to the maximum reasonable extent in those areas where redundancy could be readily applied while relying on simplex elements where redundancy would have resulted in undue complexity or other significant penalties. This philosophy results in significant differences in the extent to which redundancy is applied not only within the various portions of the guidance and control system but also throughout the total Saturn launch vehicle.

In the stabilized platform and control systems, the unreliability remains high compared to the digital system because each contains major simplex electromechanical elements. However, compared to other major systems in the launch vehicle, where little or no redundancy is incorporated, the systems look very favorable. The overall guidance and control system reliability is considered acceptable for the Apollo mission. The continuing investigations of backup schemes and design modifications being considered are merely to enhance the reliability further.

It should be reiterated that the unreliability numbers shown represent the predicted number of component or subsystem malfunctions in a million flights and not the number of mission failures. The latter, sometimes referred to as the criticality number, is derived by considering the individual failure modes and corresponding effects. The criticality numbers for the various subsystems are considerably lower than the unreliability numbers quoted.

Note that several subsystems closely related to the guidance and control system are not included in the analysis. Principal examples are the vehicle primary power source, the auxiliary propulsion system propellant source, the switch selectors which provide vehicle sequencing, and the digital command system. While these elements support the guidance and control system, they also perform other vehicle functions and support other major subsystems.

The benefits of redundancy must be traded off against the resulting penalties in weight, power, cost, and operational complexity; but the application of redundancy cannot be utilized as a substitute which permits relaxation of basic reliability design principles. High reliability component parts programs and tight quality controls must be maintained; to derive practical benefits, redundancy must be applied to a basically highly reliable system.

The Saturn guidance and control system is an inherently reliable system because major emphasis has been placed on design conservatism and simplicity, use of carefully selected component parts, and extensive testing. In addition, through judicious application of redundancy, the overall result is a system of very high reliability and flexibility. The dependability of the system has been demonstrated through three successful Saturn IB flights without a functional failure and many thousands of hours of ground testing. Weighing the results against the problems and disadvantages, we conclude that the design approach is justified and has been verified to be basically sound.

Acknowledgment

The authors express their appreciation to the various personnel within the Guidance and Control Division and the System's Engineering Office of Astrionics Laboratory and to the reliability analysts at IBM-Huntsville and IBM-Owego for their contributions. Special thanks are due Mr. John Gonsowski of IBM-Huntsville for his advice and assistance in the preparation of this paper.

Bibliography

1. Haeussermann, Dr. Walter; and Duncan, Dr. Robert C.: Status of Guidance and Control Methods, Instrumentation, and Techniques as Applied in the Apollo Project. Presentation to the Advisory Group for Aeronautical Research and Development, North Atlantic Treaty Organization, Duesseldorf, Germany, October 21-22, 1964.

2. Decher, Dr. Rudolph: The Astrionics System of Saturn Launch Vehicles. NASA TM X-53384, February 1, 1966.

3. Navigation, Guidance, and Control System Description, IBM Technical Manual, no. 66-966-0003, MSFC No. III-5-509-2, April 15, 1966 (changed May 1, 1966).

4. Thomason, Herman E.: A General Description of the ST-124-M Inertial Platform System. R-ASTR-G, NASA TN D-2983, December 1965.

5. White, Paul R.: ST-124-M 250VA AC Power Supply. R-ASTR-G-WP-10-65, April 19, 1965.

6. Kalange, M. A.; and Alcott, R. J.: Saturn V S-IC Stage Engine Gimbal Actuation System, R-ASTR-NF, May 18, 1965.

7. McGillen, V. W.; and Jacobs, M. R.: The Saturn S-II Stage Engine Actuation System. Space and Information Systems Division, North American Aviation, Inc., 650305, Undated.

8. Hamilton, M. J.: Design of a Hydraulic Gimbal System for a Moon Mission Booster Stage. Douglas Missile and Space Systems Division, Huntington Beach, California, Douglas Paper no. 3110, May 1965.

9. Control-EDS Rate Gyro Package, IBM Technical Manual, no. 66-966-0013, MSFC no. III-5-510-13, March 1, 1966.

10. Control Signal Processor. IBM Technical Manual, no. 66-966-0014, MSFC no. III-5-510-14, March 1, 1966.

11. Flight Control Computer. IBM Technical Manual, no. 66-99-0015, MSFC no. III-5-510-15, March 15, 1966.

12. Failure Effect and Reliability Analysis ST-124-M Three Gimbal Platform for Saturn C-5, R. E. File no. 7346, Bendix Corp., Teterboro, N.J., May 9, 1965.

13. Failure Mode, Failure Effect, and Criticality Analysis for S-IU-203, IBM no. 66-383-0004, MSFC no. III-2-220-4, April 1, 1966.

14. Saturn V Thrust Vector Control Failure Analysis. Boeing Document no. D5-11390, August 1964.

15. Le Roy, G. B.: Preliminary Reliability Analysis and Prediction, Moog Model-17-200A, Mechanical Feedback Servoactuator. Moog Report no. MR 975, September 11, 1964.

16. Reliability Report of Servoactuator, Mechanical Feedback (P/N 301100), Revision A. Hydraulic Research and Manufacturing Company, Burbank, California, June 8, 1966.

17. Shimizu, A.; and Schleich, W. T.: Stability and Criticality Analysis of the S-II Flight Control System. Space and Information Systems Division, North American Aviation, Inc. Report no. SID 64-1594, accession no. 66174-64, August 31, 1964

18. Reliability Engineering Model Saturn S-IVB-501 Stage. Vol. II, Failure Effect Analysis, Douglas Report SM-47406, Douglas Missile and Space Systems Division, Huntington Beach, Calif., March 1966.

19. Reliability Assessment Report for Saturn V Guidance Computer and Data Adapter. IBM Document no. 65-394-061, IBM Owego, New York, October 18, 1965.

20. Gonsowski, J. A.: Utilization of Monte Carlo Technique for TMR Reliability Predictions. IBM Document no. 63-394-059, IBM Owego, New York, Aug. 15, 1963.

21. White, J. B.: Improvement of Computer System Reliability Through Redundancy. MSFC Technical Publication, MTP-G&C-N-61-36, August 30, 1961.

No. 67-554

THE RELIABILITY CONTRIBUTION OF THE PILOT TO A LARGE LAUNCH VEHICLE CONTROL SYSTEM

by

GORDON H. HARDY and RICHARD L. KURKOWSKI
NASA Ames Research Center
Moffett Field, California
and
GLEN D. RITTER
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-554

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.04, 8.07, 12.11

THE RELIABILITY CONTRIBUTION OF THE PILOT TO A LARGE LAUNCH VEHICLE CONTROL SYSTEM

Gordon H. Hardy and Richard L. Kurkowski
Research Scientists
Ames Research Center, NASA, Moffett Field, California 94035

and

Glen D. Ritter
Research Scientist
Marshall Space Flight Center, NASA, Huntsville, Alabama 35812

ABSTRACT

An investigation was made of the reliability contribution of manual backup control to the first stage of the Saturn V launch vehicle. A technique to measure the manual control system reliability using a piloted simulator was developed. Two manual backup control systems were considered: a "Load Relief" and a "No Load Relief" system. Both systems allowed the pilot to close an adaptive control loop. This loop is parallel to the primary automatic control system. The system failure modes associated with the primary system as well as those associated with the additional hardware for the piloted backup system were considered. An analog piloted simulation was used and included rigid-body, engine-actuator, vehicle-bending, propellant-sloshing, and control-system dynamics. Nearly a thousand simulated flights with randomly selected failures were made with three test pilot subjects. The results indicate that for the failure modes and automatic system considered, the piloted manual backup system can reduce the probability of mission failure by a factor of 2. In addition, trajectory dispersions at first-stage cutoff were significantly reduced.

INTRODUCTION

Since early 1963 the Ames Research Center and the George C. Marshall Space Flight Center (MSFC) have been jointly investigating the problem of piloted guidance and control of large launch vehicles. Early works of this group and others[1-5] have established the feasibility for this class of vehicles.

At the conclusion of these feasibility studies the question was posed, "Can the probability of mission success be improved by incorporating the pilot in a backup mode to the primary automatic control system of the Saturn V launch vehicle?" The purpose of the present investigation was to answer this question. A major portion of the study was devoted to developing a technique of measuring the reliability contributed by the addition of a manual backup control system.

SYSTEM DESCRIPTION

Controlling the launch vehicle is most difficult[1-5] during the first stage of flight because of the high dynamic pressure and wind disturbances. Consequently, the investigation was limited to the first, the SIC, stage of flight.

Vehicle Description

The Saturn V vehicle configuration is shown in Fig. 1. It consists of three booster stages, an instrument unit, and the Apollo spacecraft. Fully fueled, the vehicle weighs approximately 2.7×10^6 kg. The first stage is powered by five F-1 engines. The instrument unit contains an inertial navigation and guidance system for use during boost flight. A control-system computer and necessary sensors are also located in the launch vehicle.

Trajectory

The first stage burns for approximately 150 seconds. Figure 2 shows some typical trajectory parameters. Maximum dynamic pressure, q, occurs at about 78 seconds. The maximum thrust to weight ratio, T/mg_o, reaches almost 5.

Wind Environment

The wind is the primary external disturbance during first-stage flight. Two synthetic wind magnitude profiles were used for this study. They were obtained from Ref. 6 and are shown in Fig. 3. The steady-state value of the larger magnitude profile will not be exceeded 95% of the time during the windiest month of the year nor will its vertical shear be exceeded 99% of the same time period. The steady-state value of the other profile will not be exceeded 50% of the same time period. Peak wind shear occurs near the altitude corresponding to vehicle maximum dynamic pressure. A preliminary investigation showed that the small amplitude gusts discussed in Ref. 6 had little effect on the manual control problem. Two wind directions were chosen, 135° and 225°, relative to vehicle launch heading. Previous experience had shown that quartering winds were the most difficult for piloted control.

Proposed Control Systems

Attitude of the Saturn V during the first stage is controlled by swiveling the four outboard F-1 engines. Each engine is swiveled in the pitch and yaw planes by separate hydraulic actuators. Roll is controlled by combined use of pitch and yaw actuators.

The feasibility study[1] had indicated that a piloted control system including load relief (reduction of aerodynamic loads) had merit. However, since the load relief feature requires additional hardware, two manual backup control systems were proposed for study: a "Load Relief" system and a "No Load Relief" system. Figure 4 shows the elements of these systems. The solid lines in the lower half of the figure indicate the elements of

the launch-vehicle automatic control system. Reference 7 presents a comprehensive discussion of this system. The engine actuator command signals are attitude rate and attitude error, summed, gained, and filtered in the control computer. The solid lines in the upper half of the figure indicate pertinent (to manual launch control) existing spacecraft control-system components. Existing pilot display items include the attitude, altitude, velocity, flight-path angle, etc., of the spacecraft. The dashed lines indicate the items added for the proposed manual backup systems. Both systems allowed the pilot to form an adaptive parallel control loop which he could activate when failures occurred in the primary system.

"Load Relief" System. Attitude error, from the launch-vehicle guidance system, as well as outputs from body-mounted accelerometers in the launch vehicle was added to the pilot's display for the "Load Relief" system. These accelerometers were located near the vehicle's instantaneous center of rotation,[1] such that their outputs were nearly proportional to q_α, the product of dynamic pressure and angle of attack. Aerodynamic loads on the vehicle are directly related to this product. The output of the pilot's controller was passively filtered and summed with the output of the launch-vehicle automatic system at the control computer. This filter was a passive second-order network with natural frequency of 2.7 rps and a damping ratio of 0.5.

"No Load Relief" System. The "No Load Relief" system was identical to the "Load Relief" system except that it had no body-mounted accelerometers or associated display.

Guidance and Control Constraints

The principal constraints on the launch-vehicle guidance and control system are guidance accuracy and structural loads. Since the study reported here considers the first stage of flight, structural loads were the primary constraint. The equation used for calculating the ratio of maximum vehicle structural bending moment to breakup bending moment is:

$$M = \frac{\partial M}{\partial \beta} \sum_{i=1}^{4} \beta_i + \frac{\partial M}{\partial \alpha} \alpha + \sum_{j=1}^{2} \frac{\partial M}{\partial \ddot{\eta}_j} \ddot{\eta}_j + \sum_{k=1}^{3} \frac{\partial M}{\partial \xi_k} \xi_k$$

where

M body bending moment normalized to unity at a factor of safety of one

β_i swivel angle of the ith control engine, deg

α aerodynamic angle of attack, deg

$\ddot{\eta}_j$ acceleration at nose of the jth flexible body normal mode, mps^2

ξ_k amplitude of the kth propellant tank sloshing mass, m

The effect of propellant sloshing damping forces were neglected. The partial derivatives above were assumed to be time varying. Typical values near the time of flight corresponding to high q are as follows:

$$4 \frac{\partial M}{\partial \beta} \ldots \frac{1}{5} \text{ per degree}$$

$$\frac{\partial M}{\partial \alpha} \ldots \frac{1}{11} \text{ per degree}$$

$$\frac{\partial M}{\partial \ddot{\eta}_1} \ldots 0.04 \text{ per } mps^2$$

$$\frac{\partial M}{\partial \xi_k} \ldots 0.2 \text{ per m}$$

Because of different vehicle loading for engine thrust out, the following equation was used for these cases:

$$M = \left(\frac{\partial M}{\partial \beta}\right)' \sum_{i=1}^{4} \frac{T_i}{T_n} (\beta_i \pm \beta_o) + \left(\frac{\partial M}{\partial \alpha}\right)' \alpha + \ldots$$

The primes on the partial derivatives indicate different values corresponding to the different longitudinal loading in the vehicle. The β_o term results from the unsymmetrical vehicle loading, while T_i/T_n is the ratio of actual thrust of the ith engine to nominal thrust.

While not as significant during first-stage control, trajectory dispersions were also calculated as a measure of guidance performance. Distance and velocity normal to the nominal trajectory at first-stage cutoff were used.

ANALYSIS METHOD

The systematic analysis technique used to assess the effect of the piloted backup system on mission success is similar to one that has been used for the automatic system. It is also similar to the "Pilot-Controller Integration for Emergency Conditions" concept[8] which was refined and applied to the X-22A V/STOL vehicle.[9]

The seven steps of the technique are shown in Fig. 5 and are discussed below.

(1) Define system: Collect the necessary information on the vehicle, systems, trajectory, mission, etc., to enable a simulation to be conducted. Define the manual control system.

(2) Define major failure modes: Predict major failure modes; define failure dynamics and obtain necessary information to simulate failure modes; obtain unreliability number (probability of occurrence) for each major failure mode.

(3) Simulate system and failure modes: Use the data gathered in steps 1 and 2 and appropriate mathematical models to develop a real-time piloted flight simulation of the vehicle and its major failure modes.

(4) Define pilot procedures: Using the flight simulation developed in step 3, conduct a systematic investigation wherein the failure modes investigated are made to occur at various times of flight with the pilot in control of the simulated vehicle. From this investigation, develop a background of information from which the crew can learn to detect and correctly identify each failure as well as to follow the correct pilot procedure in the event of a

failure. (Most of the emergency section of the pilot's handbook is written during this study phase. Also, at this time, preliminary changes to the proposed manual system can be made.)

(5) Conduct simulation with random failures: Using several subjects and a large number of simulated flights with random failures, determine the conditional probability of mission failure (effectivity) for each of the major failure modes.

(6) Calculate probability of mission failure: Using the unreliability numbers from step 2 and the effectivity numbers from step 5, calculate the failure mode criticality (effect of failure on probability of mission failure).

(7) Modify system and procedures as necessary: Analyze the results of step 6 to determine which failure modes have the greatest influence on mission failure. Redesign the system or modify the procedures developed in step 4 as necessary to reach a suitable level of "probability of mission success."

The application of these seven steps to the first stage of Saturn V follows.

Step (1) Define System

The system has been defined in the preceding System Description section.

Step (2) Define Major Failure Modes

The launch-vehicle major failure modes (as opposed to component failure modes) can be divided into three categories: (1) control-system hardware failures (sensors, wiring, etc.), (2) engine actuator failures (hard over, null, oscillating), and (3) thrust failures. The first 10 failures in Fig. 6 are the major failure modes for the launch vehicle considered in order of their assumed unreliabilities (probabilities of occurrence). The unreliability numbers are typical for Saturn V. Failure modes 11 through 19 are associated with the displays and controller, which were added for the piloted backup system. The unreliability data for the pilot's displays are not shown because, as will be seen later, no mission failures were caused by a display failure.

Step (3) Simulate System and Failure Modes

A comprehensive fixed-cab analog simulation was used. The display panel was representative of Apollo and is shown in Fig. 7. The all-attitude indicator in the center of the panel displayed vehicle attitude on the sphere. For the "Load Relief" manual system, attitude error was displayed on the auxiliary meters on the left and top of the indicator with the outputs of the body-mounted accelerometers being displayed on the flight director needles. Attitude error for the "No Load Relief" manual control system was displayed on the flight director needles (no acceleration). Attitude rates were presented on the three separate indicators as shown. The clock used is just to the left of the all-attitude indicator. The following nominal boost attitude profile around its circumference is used with the clock's sweep second hand. If it is assumed that the second hand is shown in the zero mission time position, it indicates that the nominal attitude should be 90°. One full minute later the nominal attitude is 66°. The nominal

attitude at staging (2-1/2 min.) is 23°. The failure warning lights are to the right and below the all-attitude indicator. The lower five lights indicate thrust loss on any one of the five F-1 engines. Of the remaining lights, only the one labeled "L/V Guidance" was used. This light indicates a failure in the launch-vehicle attitude reference platform. The six switches at the lower left of the panel could be used to open the three rate and three attitude loops in the automatic system. The "hand-controller" switch activated the pilots representative three-axis hand controller. A thumb button switch on the controller could be used (in parallel with the one on the instrument panel) to activate the controller.

Time-varying coefficient, linearized equations of motion were simulated to describe vehicle dynamics. These included three propellant sloshing modes, two flexible-body modes, six-degree-of-freedom rigid body dynamics, engine actuator dynamics, and the control-system shaping networks. A 400 amplifier analog computer complex with extensive function generation capability was used. Switching was used to allow selective simulation of the various system failure modes.

Step (4) Define Pilot Procedures

Initial simulation results were used to develop a comprehensive set of pilot procedures.

The pilot's primary task before a system failed was to monitor the displays. His only control inputs (Load Relief system only) were those necessary for load relief in the event of large wind-induced aerodynamic loads. He reduced the loads by closing the piloted parallel loop using the displayed output signals of the body-mounted accelerometers. Reducing these aerodynamic loads gives the vehicle a greater margin of safety in the event of a system failure.

In the event of a failure of the launch-vehicle system (i.e., failures 1-10), the pilot's "overriding" procedure was to "keep the attitude of the vehicle at the nominal value." He did this by operating as an adaptive element in the loop that paralleled the automatic flight-control system.

For hardware failures in the launch-vehicle control system (i.e., loss of platform, attitude rate, attitude signal, etc.), the pilot used information displayed from sensors located in the spacecraft to stabilize and control the vehicle attitude. Specifically, if the launch-vehicle attitude-rate loop malfunctions (i.e., failure 6 or 10), and the vehicle motions become unstable, the pilot, using the displayed-rate information (which is sensed from gyros located in the spacecraft), takes over and

stabilizes the vehicle motions. If hard-over control system failures occur the pilot removes the saturated signal by activating the appropriate switch on the display panel.

In the case of engine actuator or loss of thrust failures, the vehicle develops asymmetric rotational moments. In this case, the pilot acts as an integration-type element in that he injects trimming or bias commands to null the unbalanced or asymmetric rotational moments. The unsymmetrical loading caused by a loss of thrust can be further reduced by a small bias of the nominal attitude toward the failed engine. This induces an alleviating aerodynamic load on the vehicle.

When a single display failed, the information displayed was sufficiently redundant that the pilot was able to detect which instrument had failed and continue to fly the vehicle using the remaining displayed information. The pilot used the ground rule that two indications of a failure were necessary before initiating any control action.

Step (5) Conduct Simulation With Random Failures

The following items had to be considered for the actual conduct of the simulation: the basis for performance comparison, simulation variables, pilot training and briefing.

Basis for Performance Comparison. The principal consideration was, "Is the automatic flight-control system plus a piloted backup system more or less reliable than the automatic flight-control system taken alone?" The reliability level of the automatic flight-control system forms the reference condition, thus making it necessary to measure the reliability of the automatic system using the same flight simulation setup, same flight conditions, etc., that were used for the piloted system.

Simulation Variables. There were several variables to consider in the simulation: the number of failures (19 for the "Load Relief" system, 17 for the "No Load Relief" system, and 10 for the automatic system), the wind magnitude (2, previously described), time of failure (3 major times; before, at, and after high q), and the direction of the wind with respect to the failure (i.e., for some failures the vehicle turns into or away from the wind). From these variables, it was determined that there were 176 basic failure situations for the "Load Relief" system, 166 for the "No Load Relief" system, and 116 for the automatic system. To make the number of failures approximately proportional to the unreliability, 79 additional situations were added. Each of three pilots flew 255 simulated flights using the "Load Relief" system. Since many of the "No Load Relief" system situations were similar to the "Load Relief" situation, only one pilot was used for 92 simulated flights. A single unknown (to the pilot) failure at an unknown time occurred during each flight. Display and controller failures were deleted for the automatic system resulting in 195 simulated flights.

Pilot Training and Briefing. A pilot training period preceded the actual simulated flight series. The time required varied because of previous pilot experience but averaged about 30 hours per pilot. For the actual simulated flight series, the pilot was briefed on the wind direction and magnitude before each flight.

RESULTS AND DISCUSSION

Step (5) in the analysis procedure, discussed above, provided the necessary data for calculating failure mode effectivity (conditional probability of mission failure).

Simulation data for a typical system failure (one actuator hard over) and one wind magnitude (95%) utilizing the "Load Relief" control system are shown in Fig. 8. The times of failure are indicated along the abscissa in the 3 major time divisions. For the pre max q and max q time of flight, the direction in which the hard-over actuator turns the vehicle with respect to the wind is also shown along the abscissa. The maximum structural bending moment, normalized to the factor of safety of one value, experienced during the flight is presented on the ordinate. For this example, 3 flights (out of a 45 total) exceeded the breakup value. These occurred near the time for the maximum wind shear. They were considered unsuccessful flights and yielded an effectivity (conditional probability) of 0.045. This was calculated by noting that during the 40 second high q time period (150 sec total) there were 3 failures out of 18 flights. Since the failure mode unreliability numbers are assumed proportional to time and there were no failures during the pre and post max q time periods, this gives an effectivity of $(3/18)(40/150) = 0.045$. Simulation data for the example failure utilizing the "No Load Relief" and automatic systems are added in Fig. 9 and yield effectivity numbers of 0.322 and 0.488, respectively. For this example, both piloted systems improve system performance. Also, for this example, the "Load Relief" system significantly improves performance over the "No Load Relief" system (effectiveness reduced from 0.322 to 0.045). The effectivity numbers for the 19 failure modes considered are summarized in Fig. 10.

Step (6) in the analysis method (calculate the probability of mission failure) is presented in tabular form in Fig. 10. The 19 failure modes of Fig. 6 are in the first column with the unreliability numbers (probability of occurrence) in the second column. The effectivity numbers (conditional probability of mission failure, given occurrence of the failure mode) for the failures occurring with a 95% or a 50% wind are given in the fourth and sixth columns, respectively. The failure mode criticalities (probabilities of mission failure) are the products of the unreliability and effectivity numbers and are indicated for the two wind conditions in columns five and seven. The effectivity and criticality numbers are given for the three systems investigated. The systems are noted in column three. As an example, consider the typical failure situation previously discussed: one actuator hard over, Load Relief system, and a 95% wind. The failure mode criticality equals unreliability times effectivity or $(5450 \times 10^{-6})(0.045) = 245 \times 10^{-6}$ as indicated by the first bar of column five. The overall first-stage mission criticality is obtained by summing the individual failure criticalities and is shown at the bottom of the figure. This assumes that the probability of more than one failure mode occurring during the first stage is small. The results indicate (for the failure modes considered) that the addition of a piloted "Load Relief" backup control system reduces mission criticality by a factor of better than 2. For the 50% wind case, the No Load Relief system performs nearly as well as the

Load Relief, while for the 95% wind, adding the body-mounted accelerometers for the Load Relief system significantly improves the performance.

Step (7) of the analysis technique would feed back the results of Fig. 10 into changes in the system or pilot procedures. While this was not carried out for the present investigation, several interesting results may be obtained from Fig. 10.

(1) Certain engine actuator failures (nos. 1, 3, and 7 in Fig. 10) and loss of the attitude reference platform in the launch vehicle (no. 4) are the failures for which the pilot makes the largest contribution. For the other vehicle failure modes with significantly high unreliability numbers, the pilot contributed little if anything.

(2) Loss of roll attitude rate (no. 6), attitude signal saturate (no. 9), and attitude-rate signal saturate (no. 7) require opening the automatic control-system loops with one of the six switches in the lower left of the display panel (Fig. 7). Since the unreliability numbers for these failures are negligible, the criticality payoff is small considering the additional system complexity.

(3) The pilot could detect the occurrence of loss of launch-vehicle attitude platform (no. 4) or loss of launch-vehicle attitude signal (no. 8) by monitoring the displayed attitude-error signal from the launch vehicle. He compared this signal with the difference between vehicle attitude displayed on the attitude indicator (generated in the spacecraft) and the nominal attitude profile on the clock. Since the unreliability number associated with failure no. 8 is small, and failure no. 4 has a warning light in addition to the error signal, the criticality payoff associated with the display of launch-vehicle attitude-error signal again seems small.

Based on these considerations, it appears that the manual backup system of this study could be simplified without significant performance loss. The simpler system would include only the following elements in addition to those presently in the spacecraft.

(1) Three-axis proportional manual inputs summed with the output of the launch-vehicle automatic system control computer. These manual signals should be passively filtered to attenuate signal amplitudes at vehicle flexible body frequencies.

(2) The display of load relief information to the pilot. This could be from body-mounted accelerometers, angle-of-attack indicators, or perhaps the spacecraft digital computer.

Three additional results which may have application to other backup manual control systems are:

(1) The effect of pilot display failures had no effect on mission criticality (failures nos. 11 through 17 of Fig. 10 have zero effectivity) because sufficiently redundant information was presented and the ground rule used was that "two separate indications are necessary before assuming a failure."

(2) With sufficient practice the pilot could adapt to the failure dynamics from a strictly monitor mode as fast as from a more active mode of control. This is indicated by the fact that the mission criticality numbers in Fig. 10 for the 50% wind case for the Load Relief and No Load Relief systems are almost the same. For the No Load Relief system, the pilot is strictly a monitor and takes no control action until he detects a failure; whereas for the Load Relief system, he actively reduces structural bending moments due to aerodynamic loading by controlling the outputs of the body-mounted accelerometers. (The small difference in effectivity also indicates that the payoff for accelerometer control for this wind magnitude is small.)

(3) An implicit result of this study as well as earlier feasibility studies is the ease with which the pilot can filter the flexible body effects from displayed signals. In the backup mode of operation considered, the pilot was required to act as an adaptive parallel loop to the automatic system. Depending on the failure mode, he was required to close attitude, attitude rate, and/or accelerometer load-relief loops. By observing his display panel, the pilot quite easily distinguished and disregarded the flexible body content of these sensor signals; in most instances, this allowed the pilot to control rigid-body motions more effectively than an automatic system utilizing passive filtering of the flexible-body signals.

The discussion so far has centered around the maximum structural bending moment as a measure of mission success. Additional performance indices which can be considered are the trajectory dispersions. While they do not affect mission success directly in first-stage flight, they must be kept within limits to allow the upper-stage guidance system to perform adequately. The trajectory dispersions obtained for the Load Relief system are summarized in Fig. 11. Data for the automatic system flights that were successful, piloted flights for the same failure situations, and all successful piloted flights are shown for each failure. The pilot's largest contribution is when large attitude errors (i.e., thrust loss and engine actuator failures) are introduced. Acting as an integration element in the parallel piloted loop, he minimized these attitude errors. The average dispersions for all cases were reduced from 5000 to 2570 m and 91 to 47 m/sec. This result could be significant in the upper-stage guidance problem. The trajectory dispersions were also significantly reduced for the No Load Relief system (data not shown).

SUMMARY

A detailed study has been made of a manual backup control system for the first stage of Saturn V. An analysis method was developed which allowed the systematic determination of the reliability contribution of a pilot to a complex control system. The results of the analytical technique indicate that for the failure modes and automatic system considered, the piloted manual backup system can reduce the probability of mission failure by a factor of 2.

Single failures of the pilot's display instruments do not affect mission success. The

pilot was able to control a failure situation from a strictly monitor mode as well as from a more active control mode. For the Saturn V vehicle, the pilot can act as a highly effective, frequency selective filter. Trajectory dispersions were significantly reduced for the manual control case, particularly for those failure modes causing large attitude errors to be developed by the automatic system.

REFERENCES

[1]Hardy, G. H.; West, J. V.; and Gunderson, R. W.: Evaluation of Pilot's Ability to Stabilize a Flexible Launch Vehicle During First-Stage Boost. NASA TN D-2807, 1965.

[2]Dragseth, G. K.: Feasibility of Piloted Boost Control. The Boeing Co., 1962; Paper presented at TBC Symposium, March 1962.

[3]Anon.: Evaluation of Pilot Manual Control During Boost Flight. ER 11921, The Martin Co., Nov. 1961.

[4]Anon.: Titan III. Flight Control System Studies of Human Pilot Capability. Appendix I, ER 12378, The Martin Co., April 1962.

[5]Holleman, Euclid C.; Armstrong, Neil A.; and Andrews, William H.: Utilization of the Pilot in the Launch and Injection of a Multistage Orbital Vehicle. IAS paper 60-16, 1960.

[6]Daniels, G. E., ed.: Terrestrial Environment (Climatic) Criteria Guidelines for Use in Space Vehicle Development. NASA TM X-53023, 1964.

[7]Haeussermann, Walter: Guidance and Control of Saturn Launch Vehicles. AIAA paper 65-304, 1965.

[8]Cole, G., et al.: Study of Pilot-Controller Integration for Emergency Conditions. RTD-TDR-63-4092, Minneapolis-Honeywell Regulator Co., 1964.

[9]Gaul, J. W.: Application of Pilot-Controller Integration Techniques to a Representative V/STOL Aircraft. AFFDL-TR-65-200, Rep. 2226-903001, Bell Aerosystems Co., Oct. 1965.

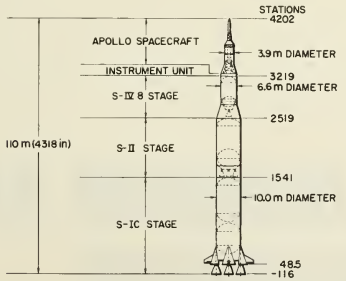


Fig. 1.- Saturn V vehicle configuration.

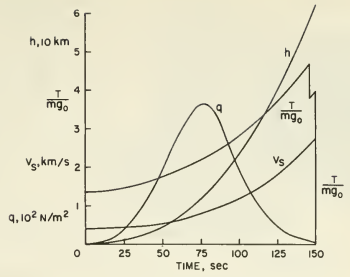


Fig. 2.- Typical trajectory data.

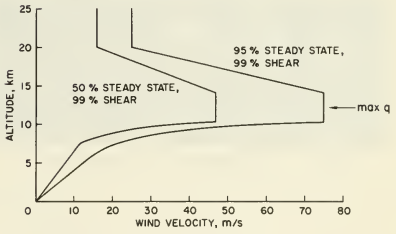


Fig. 3.- Wind profiles.

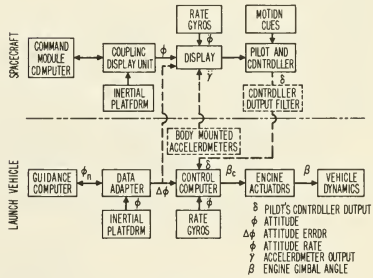


Fig. 4.- Saturn V backup guidance and control system.

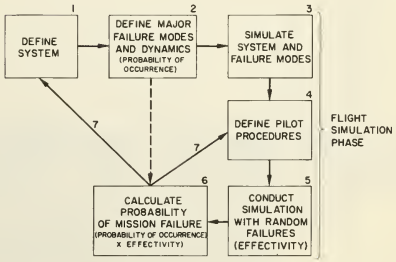


Fig. 5.- Technique for measuring probability of mission success.

67-554

NUMBER	FAILURE	UNRELIABILITY NO
1	ONE ACTUATOR HARD OVER	5450×10^{-6}
2	LOSS OF THRUST (ONE)	5000
3	TWO ACTUATORS INOPERATIVE	3100
4	LOSS OF PLATFORM	1500
5	ONE ACUATOR OSCILLATORY	1200
6	LOSS OF ATTITUDE RATE	300
7	ONE ACTUATOR INOPERATIVE	200
8	LOSS OF ATTITUDE SIGNAL	*
9	ATTITUDE SIGNAL SATURATE	*
10	ATTITUDE RATE SATURATE	*
11	ATTITUDE DISPLAY (LOCK, JUMP, DRIFT)	-
12	ATTITUDE ERROR DISPLAY NULL	-
13	ATTITUDE ERROR DISPLAY SATURATE	-
14	ATTITUDE RATE DISPLAY NULL	-
15	ATTITUDE RATE DISPLAY SATURATE	-
16	ACCELEROMETER DISPLAY NULL	-
17	ACCELEROMETER DISPLAY SATURATE	-
18	HAND CONTROLLER NULL	50
19	HAND CONTROLLER SATURATE	50

*LESS THAN 20

Fig. 6.- List of failures studied and probability of occurrence.

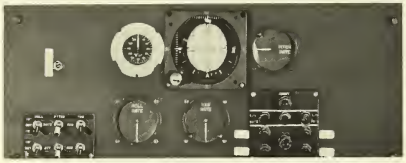


Fig. 7.- Photograph of display panel.

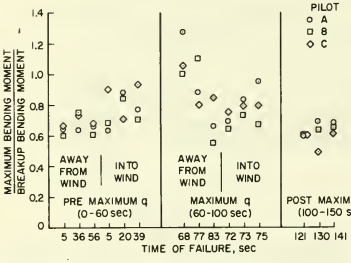


Fig. 8.- Typical system failure data using the Load Relief control system.

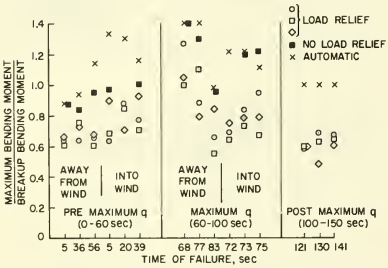


Fig. 9.- Typical system failure data for Load Relief, No Load Relief, and automatic control systems.

NO.	FAILURE	UNRELIABILITY NO.	SYSTEM	95% WIND EFFECTIVITY	95% WIND CRITICALITY 1000 2000 $\times10^{-6}$	50% WIND EFFECTIVITY	50% WIND CRITICALITY 1000 2000 $\times10^{-6}$
1	ONE ACTUATOR HARD OVER	5450×10^{-6}	LOAD RELIEF	588		0	
			NO LOAD RELIEF	588		0	
			AUTO	588		0	
2	LOSS OF THRUST (ONE)	5000	*	416		498	
				416		498	
				583		583	
3	TWO ACTUATORS INOPERATIVE	3100	*	044		055	
				185		215	
				044		044	
4	LOSS OF PLATFORM	1500	*	044		044	
				185		215	
				044		044	
5	ONE ACTUATOR OSCILLATORY	1200	*	117		282	
				490		577	
				490		577	
6	LOSS OF ATTITUDE RATE	300	*	040		0	
				001		001	
7	ONE ACTUATOR INOPERATIVE	200	*	0		0	
				0		0	
				001		001	
8	LOSS OF ATTITUDE SIGNAL	*	*	0		0	
				N.A.		N.A.	
9	ATTITUDE SIGNAL SATURATE	*	*	444		444	
				444		444	
				444		444	
10	ATTITUDE RATE SATURATE	*	*	163		185	
				1060		1200	
				0		0	
11	ATTITUDE DISPLAY (LOCK, JUMP, DRIFT)	—	*	N.A.		N.A.	
				N.A.		N.A.	
12	ATTITUDE ERROR DISPLAY NULL	—	*	0		0	
				N.A.		N.A.	
13	ATTITUDE ERROR DISPLAY SATURATE	—	*	0		0	
				N.A.		N.A.	
14	ATTITUDE RATE DISPLAY NULL	—	*	0		0	
				N.A.		N.A.	
15	ATTITUDE RATE DISPLAY SATURATE	—	*	0		0	
				N.A.		N.A.	
16	ACCELEROMETER DISPLAY NULL	—	*	0		0	
				N.A.		N.A.	
17	ACCELEROMETER DISPLAY SATURATE	—	*	0		0	
				N.A.		N.A.	
18	HAND CONTROLLER NULL	50	*	0		0	
				N.A.		N.A.	
19	HAND CONTROLLER SATURATE	50	*	240		276	
				N.A.		N.A.	
	TOTALS		LOAD RELIEF	3300×10^{-6}		2990×10^{-6}	
			NO LOAD RELIEF	3320		3320	
			AUTO	7940		6360	

* LESS THAN 20
N.A. NOT APPLICABLE

Fig. 10.- Saturn SIC criticality study; summary of results.

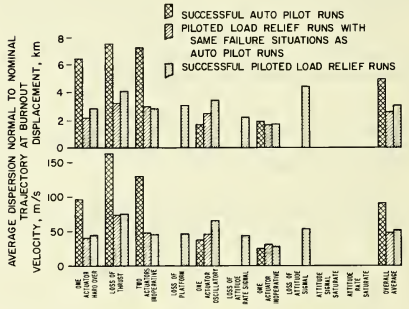


Fig. 11.- Summary of trajectory dispersion data.

No. 67-555

EFFECTIVE GUIDANCE AND CONTROL REDUNDANCY

by

B. T. BACHOFER and F. J. SPOLLEN
General Electric Company
Valley Forge, Pennsylvania
and
D. H. BARNHILL
Honeywell, Inc.
Minneapolis, Minnesota

AIAA Paper
No. 67-555

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

2.03, 10.10

-- NOTES --

EFFECTIVE GUIDANCE AND CONTROL REDUNDANCY*

B. T. Bachofer and F. J. Spollen
Voyager Spacecraft System Project
General Electric Company
Valley Forge, Pennsylvania

D. H. Barnhill
Voyager Spacecraft System Project
Honeywell, Inc.
Minneapolis, Minnesota

Abstract

The advent of multiphase, multimission programs, such as Voyager, require consideration of all possible mission outcomes in order to effectively use the weight or power allotted to improve system reliability through redundancy. The development and application of a new optimization index called Mission Expected Worth is described. By this method, values are assigned to all possible mission outcomes; each value is then multiplied by the probability of achieving that outcome; the summation of these products forms the Mission Expected Worth index. Function Flow diagrams define the mission and the single string spacecraft from which mission outcomes and values can be postulated on a "mission outcome tree". Guidance and Control components illustrate the Math Modeling of the hardware, which considers the probability of occurrence of all states that a subsystem can attain, including both normal and degraded operation of its components. Component operating states and resulting mission outcomes are related by means of a map matrix. The expected worth for each candidate configuration is computed and applied to the Optimization Computer Program which provides an ordered listing of the optimum configurations for the allotted weight. Typical input and output data are shown, and the computer systems are discussed. Current results and conclusions are discussed.

1.0 Introduction

Decisions regarding the use of redundancy in the design of a spacecraft (as well as its constituent subsystems) historically has required the engineer to assess the probability that his system, or subsystem, will perform within specifications throughout the mission life. His approach has typically been to "boil down" such reliability tradeoffs to deciding whether the change in some measurement index, such as the Reliability Figure of Merit, is worth the application of some additional resource, such as weight, power, etc. He then makes his decisions having the benefit of knowing what various alternatives will "cost" him.

The use of an index such as the Reliability Figure of Merit** for such decisions implies that it is a suitable measure of mission success. In complex spacecraft, partial or total failure of some hardware elements may still permit modes of operation which lead to valuable missions, even though they would have to be classified as partial successes. In fact, for multiphase, multipurpose missions, such as Voyager, the likelihood of a partial success (where there is a wide range between total success and complete failure) is generally greater than that of total success or total failure. As missions, and the spacecraft which perform them, become more complex, a yardstick based on more than simple success or failure is necessary to provide the engineer with the information required to make decisions regarding the application of redundancy in the spacecraft design.

This paper describes the application of a systematic approach which has recently been developed for evaluating the various possible spacecraft configurations in terms of "expected worth" within the usual constraint of weight, power or cost. By assigning values to the important possible mission outcomes, and by combining these values with the probability of attaining each outcome, an "expected mission worth" can be established. The "expected worth" may then be used as a yardstick to measure the relative merits of possible alternate configurations of subsystems within the spacecraft. The configuration which produces the maximum expected worth within the stated constraint can be called "optimum".

This new approach offers several significant advantages outlined below:

a. Multiphase, multimission programs can be accommodated.
b. Recognition is given for partial mission successes.
c. Consideration is given to degraded modes of hardware performance.
d. Allocation of resources between, as well as within, subsystems can be studied.
e. Values other than scientific or engineering may be included.

Although the principles to be described can be effectively applied to any large spacecraft, the techniques are currently being developed for

*This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration under Contract NAS7-100.

**Reliability Figure of Merit is defined as the probability that the system (or subsystem) will perform without failure for the required length of time in expected environment.

the Voyager '73 mission. Because of this, the Voyager mission and spacecraft are referred to throughout this paper. Examples have been selected primarily from the Guidance and Control subsystem.

The approach consists of four major steps as shown by the dotted lines in Figure 1. These are:

a. Mission Definition - The various phases of the mission are defined and detailed. The important mission outcomes are selected, and to these quantitative values are assigned.

b. System Definition - A baseline spacecraft, as well as various alternative configurations, are defined which satisfy the mission requirements.

c. Failure Modes & Mission Effects - The effect of failure within the potential configuration elements are analyzed and their effect on the mission is determined.

d. Configuration Selection - The possible configurations are analyzed via a computer system to select the "best" for a specified constraint (i.e., weight).

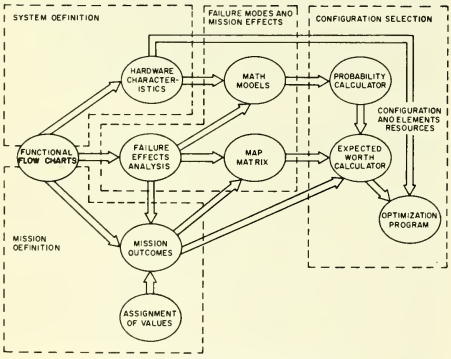


FIGURE 1 - OPTIMIZATION FLOW DIAGRAM

These steps are presented in somewhat simplified form in Sections 2.0 through 5.0, respectively. The principle result of these four steps is shown graphically in Figure 2. From the totality of all possible spacecraft configurations plotted in terms of their Mission Expected Worth (MEW) and Weight, we are interested in those which form the upper boundary of this population (shown connected by a solid line in Figure 2). For these configurations there is no better alternative, i.e., one with a higher value of MEW for the same or less weight. If desirable, a maximum value for the weight constraint may be specified beyond which no configurations will be considered.

Section 6.0 presents some results of the application of this method to Voyager. Conclusions reached thus far in the development of the method are also presented.

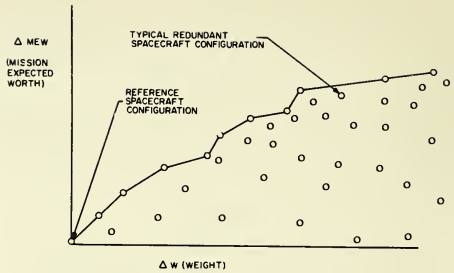


FIGURE 2. RESULTS OF THE OPTIMIZATION PROCESS

2.0 Mission Definition

2.1 Function Flow Diagrams

The Voyager mission and spacecraft system are defined via the medium of functional flow diagrams[1]. The top-level diagram of Figure 3 shows the seven mission phases for a 1973 Voyager Mission to Mars. Each of these phases is defined in progressively greater detail using lower-level functional flow diagrams.

Figure 4 is a segment of a lower level breakdown for the Transit phase. The first level breakdown can be viewed as a series of mission subphases which are performed as elements of the Transit phase beginning with acquisition of references and extending through a series of trajectory corrections with interspersed periods of cruise. Each of these first level functions can be further broken down, etc. In general, five levels are required to reach the detail where spacecraft system functions can be related directly to discrete subsystem hardware elements. In this example sensing position information, which requires a sun sensor, is shown as one of the necessary elements in accomplishing the first level function of "Acquire Celestial References". Thus, the functional flow definition serves two purposes. The higher level breakdowns provide detailed descriptions of the mission subphases. The lower level breakdowns, carried sufficiently for, can be identified with discrete hardware elements which the spacecraft must contain. This latter use of these diagrams is discussed further in section 3.1.

2.2 Mission Outcomes

In order to provide a rationale for deriving the value of a Voyager mission, various mission outcomes are postulated. By selecting the more important mission outcomes from the infinite number of possibilities and assigning specific values to them, it becomes practical to evaluate a given spacecraft configuration in terms of its expected mission worth; i.e., the sum of the products of the probabilities of all the various

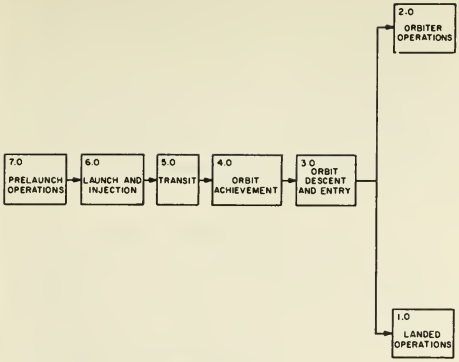


FIGURE 3. VOYAGER MISSION

outcomes and the value of those outcomes. Mission Expected Worth, then, is used as the optimization index in selecting the best allocation of redundancy within the spacecraft system.

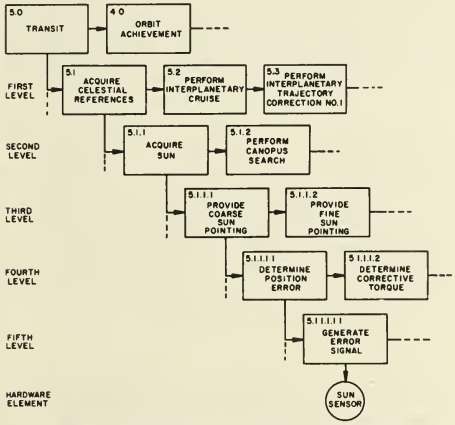


FIGURE 4. TYPICAL LOWER LEVEL FUNCTIONAL FLOW BREAKDOWN

The description of the important mission outcomes makes use of the functional flow definition that was used to describe the required spacecraft system functions throughout the various mission phases. A "tree" diagram is used to interconnect these first level functions. Figure 5 shows a simplified portion of the outcome tree developed for 1973 Voyager Mission. Across the top, the functions to be accomplished during each mission subphase are shown. In each subphase, branches of the tree represent whether that function is good, degraded or failed. Horizontal branches represent "good" states; branches with negative slopes represent either degraded or failed states. In the example of Figure 5, only good and failed branches are used. When failure or degraded performance of functions leads to missions from which negligible additional value would be accrued, the branch is terminated. If failure or degradation of a particular function still permits some additional value to be obtained from subsequent subphases, the tree is continued. In example of Figure 5, this has been done following failure of the trajectory correction functions. Following any path on the Outcome Tree from the left to any terminal node (from which there are no branches) represents a possible mission outcome.

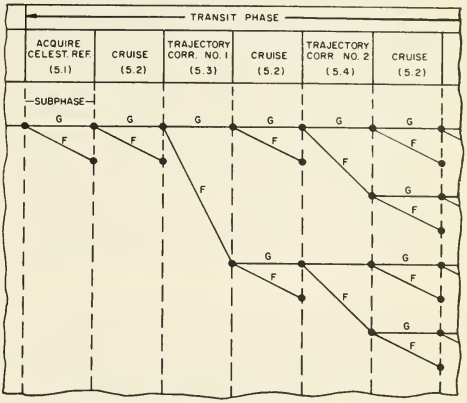


FIGURE 5-MISSION OUTCOME TREE SEGMENT

A technique has been employed in the mission outcome tree for Voyager to permit a fine gradation of outcomes while maintaining a relatively simple tree structure. This technique involves the use of "quality" states which relate to those functions associated with collecting and transmitting scientific and engineering data. During any mission subphase, value is gained through the receipt of data. Thus, the operating state of the quality function is monitored throughout the mission and modifies the potential value which would be obtained from an otherwise nominal mission. Application of this technique is discussed in the next section.

2.3 Assignment of Values

The assignment of quantitative values to mission outcomes is necessarily subjective, and there appears to be no deterministic method of calculating them. Although rigorous approaches may lead to mathematical equations, there will always be some factor or constant in the expression which depends on human judgement.

For the Nominal 1973 Voyager Mission, only scientific and engineering values were considered. For assignment of scientific value, the gross objectives of the Voyager Mars Mission were broken down into detailed subobjectives and values assigned to each according to its priority. Then, each subobjective was related to a specific instrument

in the 1973 orbiter, entry or lander scientific payload. By noting how closely the instrument payload meets its objectives, a scientific value assessment of a typical Voyager 1973 payload was performed. For assignment of engineering value, a set of rules was defined to assist in making uniform value judgements. Using these rules, engineering value assignments were made to each appropriate first level functions defined in the Nominal 1973 Voyager Mission. After the scientific and engineering value assignments were made, they were submitted to scientists, engineers and project personnel for their criticism. After several iterations, the final value structure was established.

Having arrived at quantitative values for the various mission functions, the values were allocated to the mission subphases in which they are performed, and fractional values were determined for degraded mission functions. These values were then assigned to the appropriate branches of the Mission Outcome Tree. To illustrate this, suppose that 20 value points had been assigned to Function 5.2, Cruise (Refer to Figure 5). Since this function is performed during several periods -- say 4 points for the first, 5 points for the second, etc. Now suppose the first trajectory correction were to fail. Is the following period of cruise still worth 5 points if it succeeds? If it is, we also assign 5 value points to that branch. If not, we determine what fractional value should be assigned.

It was pointed out in the preceding section that functions associated with obtaining and transmitting data may be considered as "quality" functions which modify the potential value of the in-line functions. In the above example, the 4 points associated with performing the first period of cruise should be credited only if the data obtained from such operations is received on earth. Partial receipt of this data, resulting from less than perfect performance of this quality, should result in partial realization of the 4 value points.

Suppose that we defined four mutually exclusive and exhaustive operating states for this quality function of "Obtain and Transmit Data" as follows:

Good (G) - receive both engineering and scientific data.
Degraded-one (D1) - receive scientific data only.
Degraded-two (D2) - receive engineering data only.
Failed (F) - receive neither scientific or engineering data.

Further assume that for the "G" state, we should receive 100% of the value of the particular subphase, for state "D1" 75%, for state "D2" 25% and for state "F" 0%. Now let us return to our previous example in which we assumed the first cruise subphase to be potentially worth 4 value points. We would actually expect to receive "V" value points for this subphase, where

$$V = 4 \left[1 \times P(G) + .75 \times P(D1) + .25 \times P(D2) + 0 \times P(F) \right]$$

and P(G) is the probability that this quality function is in the good state, etc. Note that the expected value is $0 \leq V \leq 4$ depending upon the value of the quality modifier (the braketed term of the above equation).

The usefulness of this quality modifier is more apparent when considering mission subphases such as "Orbit Operations" where the value is significantly affected by the type of data being returned, the data rate possible, etc.

3.0 System Definition

3.1 Function Flow Diagrams (Lower Levels)

In Section 2.1, function flow diagrams were introduced as a media for the definition of the mission and the spacecraft system. It was noted that each top-level function could be expanded in increasing detail through lower level diagrams, such as Figure 4, until the resulting functions could be correlated with discrete elements of spacecraft hardware. In this study of the application of redundancy, it has been useful to initially define a baseline or reference spacecraft system which is void of any intentional redundancy -- in other words, a system in which every element of hardware performs at least one necessary mission function. This fundamental, albeit fictitious system is termed the "single-string" system, and is the reference system to which redundancy may be added. As the single-string function flow diagrams are developed and functions are broken down into subfunctions of increasing detail, alternate methods of performing functions are encountered. At each such point, however, a choice is made to select the "best" method for the single-string system.

Having defined the single-string spacecraft system, possible redundant alternatives can be proposed. For example, alternatives discarded when defining the single-string system may be examined as candidates for functional redundancy. In addition, block multichannel and other types of redundancy can be considered as possible additions, or substitutions, to the single-string system. It is from this pool of candidate redundancy that the optimum configuration is selected for a particular weight constraint.

3.2 Hardware Characteristics

Systematic calculation of Mission Expected Worth (MEW) requires all spacecraft hardware to be described such that any component failures can be related to mission outcomes. The introduction of an intermediate plateau of "Independent Assemblies" between the hardware elements and the mission outcome tree of Section 2.2 permits a two stage Failure Modes and Effects analysis relating component failures to mission outcomes. During stage one, hardware failures lead to Independent Assembly (I/A) output states; in stage two, the I/A output states lead to mission outcomes.

The concept of Independent Assemblies involves grouping single-string and potentially redundant hardware assemblies into "families" that are related to a major functional portion of a subsystem. For example, in the Guidance and Control

subsystem, the gyros and their electronics might be considered as one family, the pneumatics a second, the autopilot a third, etc. Each family is comprised of one or more Independent Assemblies -- the single-string I/A plus any alternatives deemed appropriate -- each of which is capable of performing that function. For example, two members of the gyro family might be:

a. A package of three single axis gyros(the single-string I/A).

b. A configuration of two such packages with switching in parallel redundancy (an alternate I/A).

The Independent Assembly concept can be better understood by considering the bookcase analogy of Figure 6. Here, the bookcase represents the spacecraft, and has room to hold twenty-six books or families. Of these, seven slots are available to hold the G&C subsystem. Each slot is filled by one member from each of the seven families in the G&C subsystem. The "gyro & electronics" slot can be filled by one of the volumes labeled "single-string", or "block redundant" or "twin-axis gyros", or any other I/A which would fit that slot (i.e., would perform the functions required of the single string gyro package). Similarly, any other G&C slot could be filled with one I/A from a family of I/A's capable of filling that slot. The other nineteen slots are reserved for other subsystems. The end result is a bookcase or spacecraft configuration which can perform the complete mission and which contains one and only one book or I/A from each family. Any book can be replaced by a book from the same family, giving a new bookcase configuration with a different weight and expected worth. The expected worth changes because the probability of satisfactory operation of the family changes for each assembly. The goal is to increase the worth, while keeping the weight within limits.

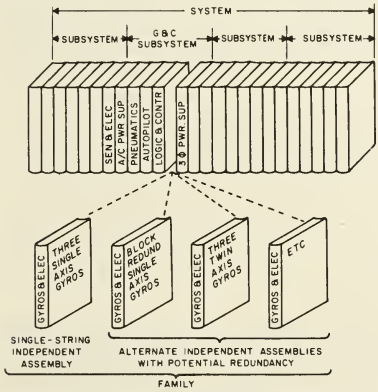


FIGURE 6-THE CONCEPT OF INDEPENDENT ASSEMBLIES

In summary then, Independent Assemblies have the following characteristics:

a. An independent assembly can be replaced in its entirety by another independent assembly from the same family of independent assemblies.

b. All functions performed by the single-string independent assembly in a family must be performed by each alternate independent assembly in the family.

c. The failure modes or states of one independent assembly are probabilistically independent of those of other independent assemblies.

d. A spacecraft configuration consists of at least one and only one member of each family of independent assemblies.

4.0 Failure Modes and Mission Effects

4.1 Failure Modes and Effects Analysis

The operating or output states of a single-string I/A are determined by consideration of the failure modes of its various components, as well as their effect on the mission. Generally, the attempt is made to limit the possible output states of an I/A to good or failed, using engineering judgement to classify degraded states, such as drift, noise, etc., as either good or failed. In some cases, an independent assembly performs multiple functions, and failure modes may exist such that a part but not all of functions can be accomplished. In this case a degraded state would be retained. For example, in the Voyager mission, the gyros and electronics are required for maneuvering (position mode) and for acquisition of celestial references (rate mode). Failures can occur which allow completion of the latter function only. This would then be classified as a degraded (D) state for that I/A.

Failure mode and effects analyses are conducted for each member of each family. The expected failure modes and their effect on the I/A output states are specified independent of the mission phase during which the failure might occur. It should be noted that the output states of interest are those of the complete I/A, not those of parts within the I/A. For example, if a I/A has dual parallel redundant components, the failure of one component still results in the good (G) I/A output state as long as the second operates satisfactorily.

4.2 Math Modeling

Each family of the spacecraft generally consists of a single-string member and one or more alternate members generally differing in the type or amount of redundancy. By definition, each member can perform the total function of the family; the probability of it doing so varies from member to member. The math model relates the probability of being in one of the possible output states derived in the failure modes analysis with the failure rate characteristics of each hardware element. This is done via a "state diagram" which shows all of the possible output states and the transition rates (failure rates) of every path between the states.

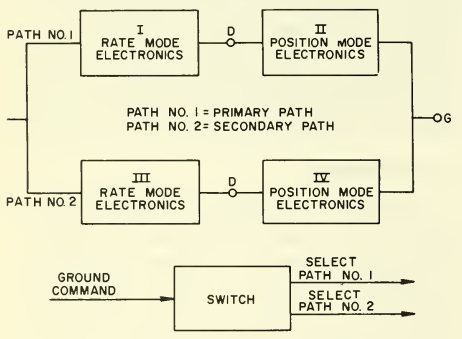
To illustrate the several steps involved in generating a state diagram[2], an I/A from the Gyro and Electronics family of the Voyager G&C subsystem will be used as an example. This family contains two gyro packages connected such that if the primary package fails, the standby package would be selected by ground command. Each package is identical to the single-string I/A so that in effect there is dual standby redundancy. The single-string I/A contains three single axis gyros, one

for each axis, with associated electronics. The gyros may operate in either a rate or a position mode, and the electronics are divided into two groups:

a. the common electronics required for the gyro to function in either mode, and
b. additional electronics which are required only during position mode operation.

Included in the latter would be torquing electronics and temperature control electronics. The good (G) state means that both position and rate mode operation is satisfactory; degraded (D) state means position mode has failed and only rate mode operation is possible; and failed (F) mode means rate mode has failed. (Since the same electronics are required for position mode operation, this mode has also failed.) To relate these modes to the Voyager mission, acquisition of references is done in the rate mode; maneuvers and velocity corrections require position mode operation where the gyros are torqued to arbitrary inertial references.

The first step in the math modeling is the generation of a Reliability Block Diagram as shown in Figure 7 for the gyro I/A. Path #1 is the equivalent of the single string case where successful operation of boxes I and II is required for the good (G) state. If the second box fails, a degraded (D) state results. The alternative shown in the Figure provides path #2 as a backup. In normal operation, path #1 would be normally selected. If failure of either box of path #1 is detected by telemetry, a ground command to the spacecraft selects path #2. Return to path #1 is allowed by means of another ground command.



OUTPUT STATES	DEFINITION
G	ACQUISITION, MANUEVER, INERTIAL HOLD CAN BE DONE
D	MANUEVER, INERTIAL HOLD CAN NOT BE DONE
F	ACQUISITION CAN NOT BE DONE

FIGURE 7-RELIABILITY BLOCK DIAGRAM

It should be noted that though various paths may be "failed" internally, the I/A has only three output states as far as the mission is concerned. Every I/A in the Gyro and Electronics family has the same three output states.

The next step is to define operating time profiles for each I/A or its components. For the gyro package, it is planned to turn off the package during the long cruise periods between earth and Mars. Other I/A's in the spacecraft will be on continuously. Where failures per million cycles is more appropriate than operating time, these profiles are generated.

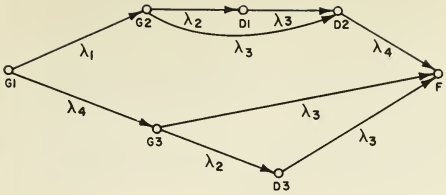
To assign quantitative estimates of failure rates to the I/A's requires the assessment of preliminary design data to identify parts which contribute to the occurrence of the failed and degraded modes. This design data can be in the form of circuit diagrams, estimated parts counts, similar hardware references, etc. Failure rate data for the various piece parts is generated from flight and test experience, analysis, and literature surveys. Based on the identified contributing parts, numerical rates for each failure mode can be established from this failure data base.

The state diagram for the gyro family example is shown in Figure 8. This is a simplified version in that the ground command switching hardware has been omitted. Each node is a possible internal state of the I/A, although the output states are still only G, D and F. Consider the top branch of the state diagram. The I/A is initially in the G1 state where both paths of Figure 7 are in the good state. If components fail such that path #1 becomes degraded (failure rate = τ_1), the ground command switches in path #2 so that the I/A state is G2. Subsequent component failures in path #2 (Failure rate = τ_2) would put the I/A in state D1 if path #2 is degraded. If path #2 should fail (τ_3), ground command would select the degraded path #1 so that D2 results. The complete failure of path #1 (τ_4) leads to the failed state of the I/A.

The lower path from G1 to F begins with the failed mode of path #1 (failure rate = τ_4) rather than the degraded mode. Ground command establishes state G3 from which failure of path #2 leads to F either directly (τ_3) or via the degraded state D3. (Note that for block redundancy $\tau_4 = \tau_1$ and $\tau_2 = \tau_3$.)

The probability of the I/A being in the good state (P(G)) is the sum of the probabilities of good states G1, G2 and G3, i.e., the probability that sufficient hardware remains after some component failure, so that the I/A can perform its function. Obviously, as more redundancy is applied, more I/A "good" states are possible, since failed portions of the I/A are replaced by redundant pieces. Each time a part fails, a new state is generated, but the I/A still remains in a "good" state.

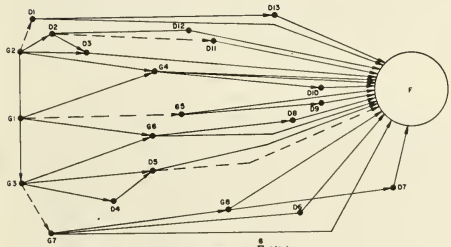
Figure 9 shows the real state diagram for this I/A after the inclusion of ground switching failure modes. Although it is obviously more complex, the entire diagram is still of a manageable size. This is not always the case, and in such instances, the math modeling is accomplished by several relatively simple state diagrams which are related to I/A output states by fairly complex logic expressions. Experience has shown this generally to be the case when there are many degraded states to be considered.



STATES	DEFINITION
GI	PATH NO. 1 AND NO. 2 ARE GOOD
G2	PATH NO. 1 IN DEGRADED STATE
DI	BOTH PATHS DEGRADED
D2	PATH NO. 2 FAILED; PATH NO. 1 DEGRADED
F	BOTH PATHS FAILED
G3	PATH NO. 1 FAILED
D3	PATH NO. 2 DEGRADED

λ_1 = FAILURE RATE FOR DEGRADED MODE OF PATH NO. 1
λ_2 = FAILURE RATE FOR DEGRADED MODE OF PATH NO. 2
λ_3 = FAILURE RATE FOR FAILED MODE OF PATH NO. 2
λ_4 = FAILURE RATE FOR FAILED MODE OF PATH NO. 1

FIGURE 8 STATE DIAGRAM SIMPLIFIED (SWITCHING NOT SHOWN)



THERE ARE 8 GOOD STATES $P(G) = \sum_{i=1}^{8} P(G_i)$

THERE ARE 13 DEGRADED STATES $P(D) = \sum_{i=1}^{13} P(D_i)$

FIG 9 ACTUAL STATE DIAGRAM OF GYRO PACKAGE EXAMPLE

4.3 Map Matrix

A "map matrix" is used to relate the family operating states defined by the mathematical model to the mission outcome tree. The map defines the logic which describes what spacecraft hardware operating states are required to perform each spacecraft function in each mission subphase. Thus, given the probabilities for the hardware states throughout the mission, the probability of the various mission outcomes may be established.

A partial map matrix for the Voyager mission is shown in Figure 10. The rows of the matrix correspond to the possible output states of various families (identified by family number); the columns of the matrix refer to the subphases of the outcome tree. As pointed out in Section 2.2, degraded subphases may be considered, as well as simple good or failed subphases. Thus, the columns under each subphase indicate the manner in which that subphase is performed. For example, the orbit insertion subphase could be performed such that the result is good (nominal orbit), degraded (non-nominal orbit) or failed (planetary orbit not achieved). The symbols at the cross points of the matrix defined at the foot of the Figure relate the outcomes to the hardware states.

INDEPENDENT ASSEMBLY FAMILY	STATE	ORBIT INSERTION 4.1			ORBIT OPERATIONS 4.2		ORBIT TRIM NO. 1 4.3		ORBIT OPERATIONS 4.4	
		G	D	F	G	F	G	F	G	F
1	G	●	●		●		●		●	
	F			●						
2	G	●	●		●		●		●	
	F			●						
3	G	●								
	D		●							
	F			●						
4	G	●	●		Ⓐ		Ⓐ		Ⓐ	
	F			●						
5	G	Ⓐ	Ⓐ		Ⓐ		Ⓐ		Ⓐ	
	DI	Ⓐ	Ⓐ							
	F			●						
6	G	Ⓑ	Ⓑ		Ⓑ		Ⓑ		Ⓑ	
	DI	Ⓑ	Ⓑ		Ⓑ		Ⓑ		Ⓑ	
	D2			●						
	F			●						
7	G						●			
	F									
8	G						●			
	F									

KEY:
- ● ALL FAMILY OUTPUT STATES ARE NECESSARY FOR THE GIVEN MISSION STATE
- ○ ANY ONE OR MORE FAMILY OUTPUT STATES ARE NECESSARY FOR THE GIVEN MISSION STATE
- ⬢ ANY FAMILY OUTPUT STATE IS A SUFFICIENT DO CONDITION FOR THE GIVEN MISSION STATE

FIGURE 10-THE MAP MATRIX

An example can best illustrate the use of this matrix.

Consider Function 4.1 (Orbit Insertion). In order to complete this function in the "good" state, the I/A's of families 1 through 4 must be in the "good" state, and 5 must be in either the "good" or "first degraded" state and 6 must be in the "good" or "first degraded" state. This could be interpreted in the form of a logic statement as follows:

$$4.1G = 1G \cap 2G \cap 3G \cap 4G \cap (5G \cup 5D1) \cap (6G \cup 6D1)$$

where 4.1G means that the result of orbit insertion function 4.1 is good, and where 1G represents the event that the output state of family 1 is good, etc. Similarly, this could be written as an arithmetic equation involving the probabilities of these events so as to arrive at the probability of completing 4.1 satisfactorily:

$$P_{4.1G} = (P_{1G})\ (P_{2G})\ (P_{3G})\ (P_{4G})\ (1 - P_{5F})$$
$$(1 - P_{6D2} - P_{6F})$$

where P_{1G} is the probability that the output state of family 1 is good in mission subphase 4.1.

The use of this type of matrix to represent the logical relationships between family output states and mission functions may, at first, appear cumbersome. It has been developed, however, so that the formulation of the arithmetic equations required to compute the probabilities associated with completing these mission oriented functions may be accomplished within the computer program. Thus, changes in the matrix, being data inputs to the computer, can be readily accommodated.

Note that the subphase 4.2 Orbit Operations requires families 1 and 2 to be in "good" state, family 6 to be in "good" or "first degraded" state, and that either Family 4 or 5 must be "good". Families 4 and 5 represent an example of functional redundancy between two different I/A's during the Orbit Operations subphase.

5.0 Configuration Selection

5.1 Probability Calculator

State Diagrams were used as models to relate the exhaustive set of I/A output states to the failure rates selected for the various hardware elements which comprise these I/A's. From these State Diagrams a set of linear differential equations may be written and solved to compute the probability of an I/A being in any output state at any time. The Probability Calculator is a computer program designed to derive these state equations and then to solve them for times corresponding to the end of each mission subphase. Since the differential equation solver uses a fourth order Runge-Kutta algorithm, high accuracy is available even for relatively large time increments. The result for each I/A is a state probability table similar to that shown in Figure 11.

	HRS	0	19.8	60.0	72.2	708.0	725.7	4788.0
	P.L.	6.0	5.1	5.2	5.3	5.2	5.4	5.2
P(G)		1.0000	.998775	.998775	.998238	.998238	.997685	.997685
P(D)		0.0000	.000056	.000056	.000099	.000099	.000143	.000143
P(F)		0.0000	.001169	.001169	.001663	.001663	.002172	.002172

WHERE:
$P(G) = P(G_1) + P(G_2) + P(G_3)$
$P(D) = P(D_1) + P(D_2) + P(D_3)$
$P(F) = 1 - P(G) - P(D)$

FIGURE 11-I/A STATE PROBABILITY TABLE

The left column of the state probability table shows the possible output states for the I/A. Across the top are references to various subphases as defined in the function flow diagrams. The time at completion of each subphase is also noted. Shown in the table is the probability of the I/A being in each state at the end of each subphase. Note that the operating time profiles used in the modeling effectively reduce the failure to zero during intervals when the hardware is not used. Thus, in the example of Figure 11, no change occurs during the subphases corresponding to interplanetary cruise (Subphase 5.2) since the gyros are turned off at that time.

The state probability tables for each I/A represent one major input to the Mission Expected Worth Calculator. This calculator, however, also requires certain subphase-to-subphase conditional probabilities. These are also provided by the Probability Calculator using the same basic techniques required to compute total probabilities.

Where satisfactory data exists, probabilities may be also put directly into the computer system without the need for mathematical modeling or use of the Probability Calculator.

5.2 Mission Expected Worth Calculator

The index for the optimization procedure is Mission Expected Worth (MEW) which consists of a combination of the probabilities of various mission outcome and the value of these outcomes. In its simplest form, MEW can be considered to be the summation of each mission outcome probability multiplied by a coefficient that represents the "value" of the associated outcome.

The basic approach used in formulating MEW consists of four steps, three of which have been described in the preceding sections. First, the more important mission outcomes were selected from the totality of possible mission outcomes. Then, specific values were assigned to these outcomes. Third, a means was established to relate the probability of achieving the various outcomes with any given spacecraft configuration via the map matrix. From these steps, the expected mission worth is computed for any given spacecraft configuration by taking the sum of the products of the probability of traversing any branch of the outcome tree times the value of traversing that branch. A value modifying term is used which depends on the operating status of certain spacecraft equipments. The first three steps are performed only once for a given mission. The fourth step is repeated in the MEW calculator each time a new configuration must be evaluated in the optimization process.

The equation to be solved by the calculator for a specified configuration of Independent Assemblies is:

$$MEW = \sum P_{ijk} Q_{ik} \Delta V_{ijk}$$

where i identifies subphase of the mission.
k identifies subphase outcomes (e.g., G,D,F)
j identifies the originating node of the subphase being considered.

The subscript "ijk" identifies the k^{th} output node of subphase i. The letter j identifies the node at which the subphase began, i.e., as a result of failed or degraded or good performance in previous subphases.

P_{ijk} is the probability of performing a given subphase function so that the system reaches node ijk.

ΔV_{ijk} is the value increment accrued by performing the subphase function and reaching node ijk.

Q_{ik} represents the quality of the data received during the subphase which modifies the value (ΔV_{ijk}) accrued during that subphase. This is equal to $a_1 q_{1ik} + ... + a_5 q_{5ik}$.

$q_{nik's}$ are the probabilities of being in a particular data quality state during subphase i.

a_n are value coefficients to modify ΔV_{ijk} depending on the quality of data obtainable.

Each value coefficient is a composite entity -- it reflects the scientific, engineering and political considerations attendant with its outcome. The actual magnitudes of the mission values are, to be sure, somewhat arbitrary and subjective. However, their relative magnitudes are readily structured on a "reasonableness" basis, and these values are treated as input data. As input data, they can be varied from one optimization run to another in order to test the results for their sensitivity to the value magnitudes.

5.3 Optimization Technique

The goal of the optimization process is to produce an "ordered listing" of possible spacecraft configurations which represent the best allocation of redundancy which can be made within the constraint specified; that is, a list of configurations having the highest Mission Expected Worth for the allotted weight. This listing is shown pictorially in Figure 2 where the weights of all the possible configurations are plotted against their MEW. The configurations having the highest MEW for any weight are shown by the upper boundary. A brute force approach of calculating MEW for all possible configurations is prohibitive because of the large number of possible configurations (about 10^{21} for the Voyager). Thus, a special technique is required to rapidly reduce the number of configurations to be considered.

The optimization approach used is based on the following assumptions:

a. The minimum increase in Mission Expected Worth (MEW) gained by using an alternate assembly in a particular family occurs when the single-string I/A is used in all other families. This minimum value will be called δ_oMEW.

b. The maximum increase in MEW gained by using an alternate assembly in a particular family occurs when all other families contain the I/A used in the MEW_{MAX} configuration. (MEW_{MAX} configuration contains the alternate in each family having the largest δ_o MEW).

c. Optimum or very nearly optimum configurations are obtained by considering alternates in the order of decreasing values of δMEW/ΔW. (ΔW is the increase in resource requirements resulting from the use of this alternate.)

The following steps summarize the salient steps of the optimization technique (refer to figure 12):

a. Divide the weight scale into equal segments between the reference configuration weight and the maximum weight value to be considered.

b. Consider families in order of decreasing value of δMEW/ΔW. Select the family with the highest $\frac{\delta MEW}{\Delta W}$ and insert each of its alternates into a configuration where all other families contain the single-string or reference alternate. The result is configurations with different MEW's and weights.

c. Discard all configurations within the same weight increment except the one with the highest MEW, i.e., save for further consideration only the configurations in each weight segment having the largest MEW value.

d. Apply each alternate of the family with the second highest $\frac{MEW}{W}$ to each of the saved configurations thus forming new configurations and again save only the one in each weight segment having the largest MEW value.

e. This method is applied until the last family has been considered.

f. The final result will be a list of not more than one configuration in each weight segment. These are considered to be the optimum configurations for any allotted weight.

Additional discard criterion are used in the actual optimization program, and there are various other options provided in the optimization method. These generally tend to reduce the computer running time at the expense of negligible loss of accuracy in the results.

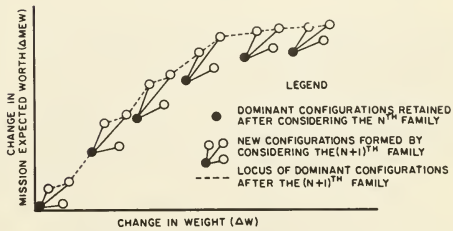


FIGURE 12. THE OPTIMIZATION PROCESS

67-555

6.0 Application To The Voyager Mission

6.1 Analysis of Voyager Mission

The preceding sections have outlined the principles involved in describing a mission and spacecraft in a format suitable for application to a computer system designed to optimize reliability through application of redundancy. The mission and spacecraft for the '73 Voyager mission to Mars were reduced to such a format, and the computer programs and procedures were applied for the first time.

One immediate observation upon applying this method is the large dimensionality of the problem. The Voyager spacecraft was divided into 59 families of independent assemblies, and a total of 174 different independent assemblies were postulated to fill the 59 available slots. Since the 59 single-string members are included in the 174, this leaves 174 - 59 = 115 alternatives, or an average of almost two alternate assemblies per family. This number in itself seems small enough until one realizes that some 10^{21} different system configurations can be formed from these 174 independent assemblies. To determine which of these system configurations are optimum within preselected bounds on system weight, cost, or power is a trade study of the first order.

The 59 families and 174 independent assemblies of the Voyager system were divided among the subsystems as shown in Table 1. The breakdown of the seven families of the Guidance and Control Subsystem is shown in Table 2. Estimates of the power and weight of each of these G&C alternates are shown in Table 3.

Portions of the mission outcome tree and map matrix were shown in Figures 5 and 10 in order to illustrate the principles under discussion. The Voyager '73 outcome tree contains 27 subphases leading to more than 200 distinct valuable outcomes. By the time that quality modifiers are added, the number of mission outcomes exceeds 10^5 The map matrix is a grid 75 subphase outputs by 128 family output states so that a total of 9560 entries can be made at the cross points. It should also be noted that both the tree and the matrix have been developed for the first level functions of the function flow diagram (see Figure 4). Any lower level would substantially multiply all the dimensions.

In the application of this technique to the Voyager '73 mission, the mission outcome tree, the map matrix, the number of I/A's and their alternates, the maximum number of output states for one I/A Family, etc. have incorporated many desirable, though possibly not necessary, features. This has been done intentionally because of the relative ease with which dimensions or options may be reduced or eliminated if experience in exercising this tool shows this to be desirable.

A portion of a typical optimization printout is shown in Figure 13. Two spacecraft configurations are shown. Each family of the spacecraft is identified by a two letter code. Directly beneath the code is a number indicating the I/A within that family which was selected. The total weight and value of MEW for that configuration is also printed out. A listing of up to 150 "best" configurations may be defined over the weight range from the reference configuration weight to any specified weight constraint.

SUBSYSTEM	NUMBER OF FAMILIES	NUMBER OF INDEPENDENT ASSEMBLIES
COMPUTER AND SEQUENCER	12	36
ENGINEERING MECHANICS	13	25
GUIDANCE AND CONTROL	7	27
POWER	8	9
PROPULSION	8	18
SCIENCE	8	8
TELECOMMUNICATIONS	3	11
TOTALS	59	174

TABLE 1 VOYAGER SPACECRAFT FAMILIES

GUIDANCE AND CONTROL FAMILY	NUMBER OF INDEPENDENT ASSEMBLIES
ATTITUDE CONTROL SENSORS AND ELECTRONICS	7
GYROS AND ELECTRONICS	4
PNEUMATICS	6
ATTITUDE ELECTRONICS POWER SUPPLY	2
AUTOPILOT ELECTRONICS	3
LOGIC CONTROL UNIT	3
GYRO PACKAGE POWER SUPPLY	2
TOTAL	27

TABLE 2 GUIDANCE AND CONTROL FAMILIES

FAMILY NAME	ALTERNATE NO.	WEIGHT (LBS)	POWER (WATTS)
A/C SENSORS AND ELECTRONICS	1	17.70	7.20
A/C SENSORS AND ELECTRONICS	2	24.30	28.20
A/C SENSORS AND ELECTRONICS	3	26.90	27.90
A/C SENSORS AND ELECTRONICS	4	21.00	15.60
A/C SENSORS AND ELECTRONICS	5	22.00	16.10
A/C SENSORS AND ELECTRONICS	6	18.40	7.90
A/C SENSORS AND ELECTRONICS	7	17.90	7.45
GYROS AND ELECTRONICS	1	7.30	10.80
GYROS AND ELECTRONICS	2	15.60	22.10
GYROS AND ELECTRONICS	3	20.60	26.60
GYROS AND ELECTRONICS	4	16.60	23.00
PNEUMATICS SUBASSEMBLY	1	108.20	0.30
PNEUMATICS SUBASSEMBLY	2	242.30	0.60
PNEUMATICS SUBASSEMBLY	3	216.30	0.60
PNEUMATICS SUBASSEMBLY	4	158.30	0.60
PNEUMATICS SUBASSEMBLY	5	109.20	0.30
PNEUMATICS SUBASSEMBLY	6	239.20	0.60
A/C ELECT POWER SUPPLY	7	1.60	4.40
A/C ELECT POWER SUPPLY	8	3.20	5.40
AUTOPILOT ELECTRONICS	1	3.00	4.40
AUTOPILOT ELECTRONICS	2	7.00	9.80
AUTOPILOT ELECTRONICS	3	5.00	6.50
LOGIC CONTROL UNIT	1	0.60	1.20
LOGIC CONTROL UNIT	2	1.80	3.60
LOGIC CONTROL UNIT	3	1.20	2.40
GYRO PACKAGE POWER SUPPLY	1	2.50	2.40
GYRO PACKAGE POWER SUPPLY	2	5.00	4.80

TABLE 3
GUIDANCE AND CONTROL SUBSYSTEM
INDEPENDENT ASSEMBLIES

FIGURE 13

FINAL CONFIGURATIONS

SC SD SE SF SG PA PB PC PD PM PN PP PU QA QB QD QE QF QG QH CA CD CE CF CG CH CJ CK CL CM CN CP EA EB EC ED EE EF EG EH

A A A A A A A A A A B A A A B A B A C A H A C B B D B A A B B A A A P R B B B B

TA TB TC MA MC MD MG MH ML MR LA LB SH MB ME MM MN MP MQ SA SB

C B B A A E A A A A A A A A A A A A A A A A

RESOURCE = 17115,22 MEW = 113,1863

SC SD SE SF SG PA PB PC PD PM PN PP PU QA QB QD QE QF QG QH CA CD CE CF CG CH CJ CK CL CM CN CP EA EB EC ED EE EF EG EH

A A A A A A A A A A B A A E B B B A C A H A C B B D B B A B B B A A P R B B A B

TA TB TC MA MC MD MG MH ML MR LA LB SH MB ME MM MN MP MQ SA SB

C B B A B E A A A B A A A A A A A A A A A A

RESOURCE = 17260,72 MEW = 133,3105

6.2 Optimization Results

There have been two types of exercises for which the redundancy optimization program has been used since its development and checkout. First, it has been used for a series of test runs to develop insight and confidence in its operation. Second, it has been applied in the normal sense to identify spacecraft configuration selections and to begin development of parametric data with respect to probability of success numbers, assignment of values, etc. Examples of some of the results obtained in the latter application are contained in the following paragraphs.

First, in order to develop correlation between MEW and the more familiar Reliability Figure of Merit (RFM), consider the graph shown in Figure 14. Here the results of two optimization runs are plotted. Each point represents the MEW value and weight corresponding to the final configurations produced by the optimization program. The upper curve represents the results obtained with value points distributed throughout the various subphases. In this case partial value is awarded for partial success. The lower curve represents the results obtained by assigning all value to completion of the total mission (assumed to be the completion of one month of planetary orbital operations). In either case it is possible to obtain 200 value points if the mission is totally successful. If we selected a spacecraft configuration with a weight of 17,350 pounds, for example, this would (from the upper curve) correspond to a MEW of 134. Now for this same configuration weight, MEW would correspond to 112 if our criteria for evaluation were total mission success. In other words a spacecraft configuration weighing 17,350 pounds would have a probability of 0.56 of successfully completing its entire mission. It also would have a Mission Expected Worth of 134 (out of 200 points distributed throughout the mission).

It may be interesting to note at this point the general shape of the MEW-Weight curves of Figure 14. Allocating additional weight initially produces configurations of substantially higher MEW values. Beyond a weight of 17,050 pounds, however, the value of MEW increases much less rapidly, reaching a value that never exceeds about 134. This might suggest concluding that it would be inefficient to allocate any more weight to the spacecraft than about 17,050 pounds. It might also suggest that it would be fruitful to identify those Families of Independent Assemblies which have a dominate role in limiting the value of MEW and consider additional alternates in these cases. The latter, in fact, is currently being considered.

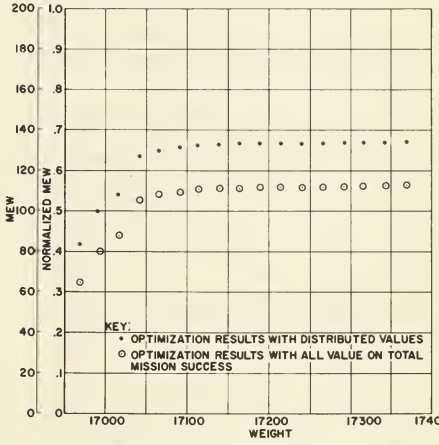


FIGURE 14. CORRELATION OF MEW WITH REM

Another series of optimization runs were made to determine the sensitivity of MEW and the configuration selections to the particular value assignments made to the mission outcome tree. Four different value assignments were made and compared to the nominal value assignments derived from a study of the mission objectives. They are shown in Table 4 and described below:

a. Line one shows the nominal value assignments used.

b. Line two places all value on the spacecraft. The value previously assigned to the capsule is assigned to the spacecraft in the Orbiter Operations mission phase.

VALUE ASSIGNMENT	LAUNCH & INJECTION			TRANSIT										ORBIT ACHIEVEMENT						ORBIT DESCENT & ENTRY			CAPSULE OPERATIONS		ORBITER OPERATIONS			TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1. NOMINAL VALUE ASSIGNMENT	1.0	1.0	4.0	1.0	.25	1.0	.5	.5	.5	.25	--	4.0	--	12.0	6.0	4.0	4.5	2.0	3.5	4.0	4.0	2.0	27.0	74.0	22.0	2.0	19.0	200.0
2. ALL VALUE ON S/C	1.0	1.0	4.0	1.0	.25	1.0	.5	.5	.5	.25	--	4.0	--	12.0	6.0	4.0	4.5	2.0	3.5	4.0	4.0	2.0	--	--	73.0	2.0	69.0	200.0
3. ALL VALUE ON CAPSULE	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	54.0	146.0	--	--	--	200.0
4. ALL VALUE ON S/C (27)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--200.0	200.0
5. ALL VALUE ON S/C (25)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	200.0	--	--	200.0

TABLE 4. VALUE ASSIGNMENTS USED IN OPTIMIZATION RUNS

c. Line three places all value on the capsule.

d. Line four assigns all value to the spacecraft in subphase 27. In this case the spacecraft must complete its total mission successfully including six months of planetary orbital operations in order to receive any value.

e. Line five assigns all value to the spacecraft in subphase 25. This corresponds to the above case except that only one month of planetary orbital operations is required.

Figure 15 shows the results of these runs. The loci of configurations rather than discrete configurations are shown for simplicity. The general sensitivity of the magnitude of MEW to these changes in value assignment is apparent. In particular, the requirement of six months of successful orbit operations drastically reduces MEW over the nominal case. From examination of the configurations selected for various weights, however, it was found that the configuration selection is much less sensitive to value changes, particularly when values are not shifted between the capsule and spacecraft systems. For example, the configurations selected for various weights are either quite similar or identical when using the value assignments of lines 1, 2, 4 or 5 of Table 4. When all value is placed on the capsule system, however, significant configuration changes do occur. These changes occur primarily in those I/A Families not required to deliver the capsule. (ie., Families associated with obtaining orbital planetary science revert to the single-string alternate.) Thus, configuration selection is somewhat insensitive to the distribution of value assignment within spacecraft system.

In the results discussed thus far, it should be noted that the optimization program arrives at the allocation of redundancy for a single spacecraft. The mission expected worth for any configuration selected is therefore based on the performance of a single spacecraft. Figure 16 shows how Mission Expected Worth is affected by using two Planetary Vehicles (P/V) per Launch Vehicle (L/V) (assuming no increase in MEW for repeated performance). Also shown is the effect of launch vehicle reliability. For the "normal L/V reliability" curves, a value of .85 was used as the probability of successful performance of the launch vehicle from lift-off through achievement of interplanetary trajectory injection. For "perfect L/V reliability" the probability was set at unity. The effective changes in MEW for various combinations can be seen. Note that the MEW values are extremely sensitive to the launch vehicle probabilities, as would be expected. The selected configurations, however, are not. In fact, the identical spacecraft configurations are selected irrespective of the L/V probabilities used.

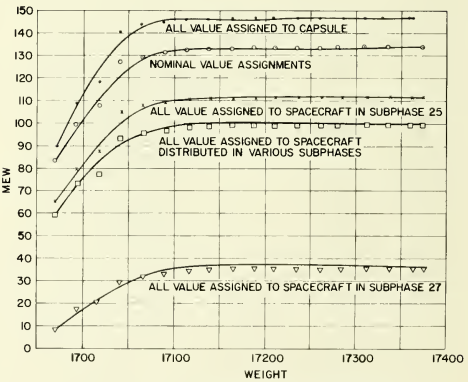


FIGURE 15 OPTIMIZATION RESULTS WITH ALTERNATE VALUE ASSIGNMENTS

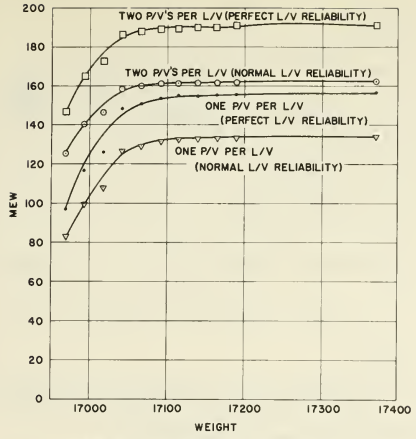


FIGURE 16 MISSION EXPECTED WORTH WITH ONE AND TWO P/V'S PER L/V

6.3 Conclusions

Based on results similar to those cited in the foregoing paragraphs, it is concluded that a workable configuration optimization tool useful in the application of redundancy has been developed. Further, this tool, although presently oriented towards the Voyager mission, is adaptable to other space flight missions.

It is important to note that in the development of this tool, decisions often had to be made as to whether or not to simplify an assemblage of facts. For example, a decision had to be made as to whether or not to consider degraded failure modes of hardware assemblies. For another example, a decision had to be made as to whether or not to consider degraded failure modes of the mission. In nearly all cases, the decision was made not to simplify. By employing this strategy, it was recognized that some facets of the tool may be unnecessarily complicated. However, this strategy does permit rapid simplifications as soon as such a need is clearly recognized.

A closely related decision was to avoid assumptions and approximations. In the case of the optimization technique itself, however, assumptions were necessary because of the general complexity of the Mission Expected Worth formulation. Thus, the optimization procedure cannot be guaranteed to always find the true optimum configurations. This aspect, however, has been independently investigated by the Information and Control Laboratory of Stanford Research Institute [3] with the general conclusion that this same difficulty would be encountered with all optimization procedures presently available, and the dynamic programming method used is currently the best approach. Experience as well as analysis both indicate that the optimization procedure will rarely yield the wrong answer and even if it does, the deviation from the true optimum is small.

It is believed that the development of this computerized system will provide the designer with a powerful tool to be used throughout the design stages of a spaceflight program.

Appendix

Definition of Terms

Mission Subphase - The seven Voyager mission phases are divided into subphases which are the first level "in-line" functions for each phase (the functions that generally define the spacecraft trajectory functions and operating modes).

Mission Outcome Tree - This is a diagram used to describe the various mission outcomes, and is constructed using first level "in-line" functions.

Quality Functions and Value Quality Co-efficient - "Quality" functions are used to incorporate into the mission outcome tree those support functions associated with collecting and transmitting scientific and engineering data. Various operating states of this "quality" function are used to modify the "potential value" which would be obtained from an otherwise nominal mission.

Single String - "Single-string" hardware elements are those necessary to perform the required spacecraft functions but void of all hardware used solely for backup or redundancy.

Family - A family is a major functional portion of a subsystem which performs its function independently of other parts of the system.

Independent Assemblies (I/A) - A family is made up of a number of independent assemblies which are the alternate configurations of that family. Each independent assembly is capable of performing the function of its family independent of other hardware.

Spacecraft Configuration - A spacecraft configuration consists of one and only one I/A from each family of I/A's.

Map Matrix - This is a matrix which shows the relationship between the operating states of each I/A family and each spacecraft function used to describe the mission outcomes.

State Diagram - A state diagram is a systematic means of representing the status of an I/A (good, degraded, fail, etc.) and the transitions from state to state.

Probability Calculator - The probability calculator is a computer program that solves the differential equations set up from the state diagrams and calculates the probability of each I/A being in a particular output state.

Mission Expected Worth (MEW) - Mission Expected Worth for a configuration is the sum of the expected worth for all mission outcomes shown as branches on the tree. The expected worth of each subphase outcome is the product of the probability, incremental value for the subphase, and quality function for each subphase outcome.

Mission Outcome Value - Mission outcome value is made up of scientific and engineering value associated with each subphase outcome.

MEW Calculator - The MEW calculator is a computer program subroutine that calculates the mission expected worth for any spacecraft configuration.

Resource Data - The resource is the weight, power consumption, cost or any other parameter which is to be constrained during the optimization procedure.

Optimization Program - The optimization computer program considers the possible spacecraft configurations and selects the configuration that has the largest MEW within a given resource segment.

Resource Segment - The Optimization Program divides the resource scale between the value for the reference configuration and the maximum resource allowed into "m" segments which are known as resource segments.

Symbols

i - This subscript identifies the mission subphase.

j - This subscript identifies the node or outcome from which subphase i begins

k - This subscript identifies the outcome node of subphase i .

$\cap$ - Logic symbol for "intersection" or "and" function.

$\cup$ - Logic symbol for "union" or "or" function.

λ - Failure rate in %/1000 hours.

References

1. Air Force Systems Command, System Engineering Management Procedures, AFSCM 375-5, December, 1964.

2. Air Force Systems Command, Practical Applications of Electromechanical Redundancy for Flight Control Systems, AFFDL-TR-66-31, August, 1966.

3. Stanford Research Institute, Evaluation of an Approach to Optimum Selection of Redundant Components with Respect to Mission Value, R. Larson, R. Ratner and D. Luenberger, May, 1967.

No. 67-556

INITIAL ALIGNMENT OF A STRAPDOWN INERTIAL REFERENCE AND NAVIGATION SYSTEM

by

HANS F. KENNEL
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-556

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

8.02, 8.04, 8.10

INITIAL ALIGNMENT OF A STRAPDOWN INERTIAL REFERENCE AND NAVIGATION SYSTEM

Hans F. Kennel
Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

The initial alignment of a specific strapdown system is treated in detail. Three single axis platforms and three pendulous integrating gyro accelerometers are the instruments for the strapdown inertial measuring unit (IMU). A coordinate transformation matrix computer of the differential analyzer type completes the system. This computer is actively utilized during the initial alignment. The direction of gravity and the line of sight to an azimuth theodolite are the inputs for the alignment. The alignment is performed on the launch pad; therefore, the inputs are corrupted by noise (wind-induced and other vibrations). Filtering of the noise is handled in an open loop fashion; i.e., the alignment data are recorded and subsequently processed by a general purpose computer (onboard or ground-based). The calculated misalignment is then eliminated almost instantaneously. The procedure must be repeated at least once because of error cross coupling. The system and its alignment will find application in the proposed MSFC strapdown IMU flight tests. If the tests show adequate accuracy, the system could be used as a backup or a replacement for the present Saturn V ST124 stabilized platform. The mathematical derivation of an orthonormalization method is appended.

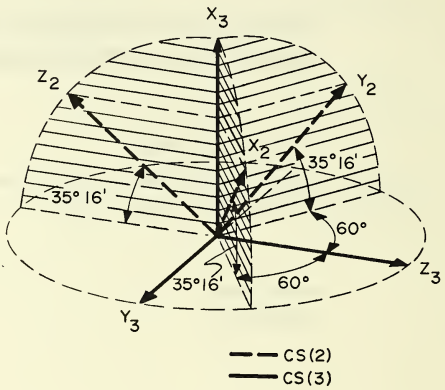
Coordinate System Definitions

The coordinate systems (CS) used in this paper are listed below. Parallel CS's are considered equivalent since only angular relationships are important for the initial alignment. The three axes of the various CS's will be called x_i, y_i, and z_i, where i denotes the CS to which the axis belongs. Only two axes will be defined; the third completes a standard right-handed CS.

CS(0) Earth-fixed launch-site CS; positive x-axis upward parallel to the local vertical as defined by a plumb bob; positive z-axis downrange.

CS(1) Strapdown inertial reference CS, rotating with earth until removal of earth rotation rate compensation at guidance release; positive x-axis ideally upward parallel to the local vertical (on launch pad); positive z-axis ideally downrange. This CS will be misaligned with its ideal orientation, which coincides with CS(0) as long as the vehicle is on the launch pad.

CS(2) Strapdown inertial instrument CS formed by the input axes of the single axis platforms and the pendulous integrating gyro accelerometers, disregarding any instrument misalignment. The orientation of CS(2) with respect to CS(3) is illustrated as follows:



CS(3) Strapdown optical cube CS, ideally (on the launch pad) aligned with the positive x-axis parallel to the local vertical and the positive z-axis parallel to a radius toward Position I (approximately 90°12' east of north) of the vehicle.

CS(4) CS of the porro prism rotation axis (the rotation axis is nominally parallel to x_3), misaligned with CS(3) by two small angles, δ_{y3} and δ_{z3}, about the y_3- and the z_3-axes. No misalignment about the x_3-axis is assumed here.

CS(5) CS rotated with respect to CS(4) about the x_4-axis by the porro prism encoder angle α_p plus a small error angle δ_{x4}. The porro prism encoder angle is nominally zero when the prism is pointed in the y_4-direction.

CS(6) CS rotated with respect to CS(5) about the y_5-axis by a small angle δ_{y5}, which is the tilt angle of the porro prism dihedral edge.

CS(7) Earth-fixed launch-site CS, rotating with earth, positive x-axis upward along the local vertical, positive z-axis due east.

Disregarding errors and misalignments yields

$$CS(0) = CS(1)$$
$$CS(3) = CS(4)$$
$$CS(5) = CS(6) .$$

If the launch azimuth is 90°12',

$$CS(0) = CS(1) = CS(3) = CS(4) \approx CS(7)$$
$$CS(5) = CS(6).$$

Nomenclature

A	any perfectly orthonormal transformation matrix
A_0	one specific matrix of all possible A's
a_c	centrifugal acceleration caused by earth rotation rate at the launch pad
a_Ω	apparent acceleration, measured by the PIGA's, caused by earth rotation rate
${}^bC_a$	orthonormal transformation from CS(a) to CS(b)
$C \equiv {}^1C_2 \equiv [c_{ij}]$	transformation matrix which is updated by the CTMC
c_{ij}	elements of C
Δc_{ij}	small changes in c_{ij} used to update C by difference equations
d_i	direction cosines of a fixed rotation axis (sometimes d_{ij} if there are more than one axis)
g_0	measured acceleration caused by gravity, positive upward, on launch pad
I	three-by-three identity matrix
L_{ij}	component of the LOS unit vector along the i-axis of CS(j)
R	orthonormal rotation matrix
t	mean solar time in seconds
α_j, β_j, and γ_j	rotation about the x-axis of CS(j), y-axis of CS(j), and z-axis of CS(j), respectively
α_p	indicated porro prism encoder angle
α_R	reference azimuth
$\alpha(t)$	time varying addition to the reference azimuth
γ_0 and γ_p	LOS depression in CS(0) and CS(6), respectively
δ_{ijk}	error angle about the i-axis of CS(j) caused by influence of K
δ_A and δ_L	azimuth error and leveling error, respectively
Δ_{ij}	elements of $[\Delta_{ij}]$, see Notation
ϵ	orthonormalization error matrix, in general
$\epsilon_0 \equiv [\epsilon_{ij}]$	orthonormalization error matrix with "best fit"
ϵ_n	ϵ of the nth iteration
ϵ_n'	ϵ' of the nth iteration
$\|\|\epsilon\|\|$	norm of ϵ
ϵ_{ij}	elements of ϵ_0
$\dot{\theta}_i$	any set of angular velocity components (used in Notation)
ϕ	rotation angle about a fixed axis (see Notation)
$\Delta\phi_i$	incremental readout of $\int \dot{\phi}_i dt$
$\dot{\phi}$	angular velocity of strapdown IMU
$\dot{\phi}_i$	components of $\dot{\phi}$ in CS(2)
ϕ_0	geodetical latitude of the launch pad
Φ	see Notation
Ω	earth rotation rate
Ω_i	components of earth rotation rate in CS(0)

Notation

$$[a_{ij}] \equiv \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

67-556

$$[\dot{\theta}_i] \equiv \begin{bmatrix} 0 & +\dot{\theta}_3 & -\dot{\theta}_2 \\ -\dot{\theta}_3 & 0 & +\dot{\theta}_1 \\ +\dot{\theta}_2 & -\dot{\theta}_1 & 0 \end{bmatrix}$$

$$\Phi \equiv \begin{bmatrix} d_1^2 v\phi + c\phi & d_1 d_2 v\phi + d_3 s\phi & d_1 d_3 v\phi - d_2 s\phi \\ d_2 d_1 v\phi - d_3 s\phi & d_2^2 v\phi + c\phi & d_2 d_3 v\phi + d_1 s\phi \\ d_3 d_1 v\phi + d_2 s\phi & d_3 d_2 v\phi - d_1 s\phi & d_3^2 v\phi + c\phi \end{bmatrix}$$

with $v\phi \equiv \text{vers}\,\phi = 1 - \cos\phi$; $c\phi \equiv \cos\phi$; $s\phi \equiv \sin\phi$

d_i = constant; $\phi = \phi(t)$.

Introduction

A space vehicle inertial reference system must be aligned with its launch-pad-fixed reference before flight. Accuracy of this alignment is very critical since a large portion of the injection errors are caused by alignment errors.

The inertial sensing instruments of the strapdown inertial reference system are "strapped down" to the vehicle frame and partake therefore in the vehicle angular rotations. Because the inertial reference coordinate system (CS) must be calculated, the strapdown system is also called an analytic platform. Whereas the initial alignment of an inertially stabilized platform is a mechanical procedure, the alignment of a strapdown reference CS is a purely mathematical operation. This mathematical operation tries to reduce, as much as possible, the attitude error of the calculated reference CS with respect to the desired guidance reference CS. This attitude error must be reduced while the vehicle is on the launch pad under the influence of wind-induced vibration and sway.

One specific alignment scheme is treated in detail rather than many possible alignment schemes in general. For such a detailed discussion it is necessary to specify the type of vehicle and inertial sensing instruments and a coordinate transformation matrix computer (CTMC). The initial alignment is treated for a vehicle of the Saturn class with a strapdown inertial measurement unit (IMU) which utilizes three single axis platforms (SAP's) and three pendulous integrating gyro accelerometers (PIGA's). The CTMC is organized somewhat like a digital differential analyzer (DDA) and used to keep the strapdown reference CS space-fixed during flight and launch-pad-fixed before flight.

The alignment scheme will be tested with the proposed flight experiment of a strapdown IMU with the previously mentioned CTMC.[1] This flight experiment will be flown within the next two years and will indicate the usefulness of a strapdown inertial reference system either as a back-up reference system or as the main reference system for Saturn V.

Basic Alignment Scheme

The alignment scheme is an iterative procedure; each iteration is broken down into a leveling and an azimuth alignment of the strapdown reference CS, which is launch-pad-fixed until guidance release, i.e., removal of the earth rotation rate compensation.

For the leveling, the outputs of the PIGA's are used to determine the local vertical direction (as defined by a plumb bob) with respect to the inertial sensing instrument triad. Alignment in azimuth is done with the aid of a line of sight (LOS) from the strapdown IMU to an azimuth theodolite. The LOS azimuth and depression (as viewed from the strapdown IMU) are accurately known in CS(0).

The PIGA outputs are corrupted by accelerations caused by sway and vibrations induced by the wind and other perturbations. The local vertical must be recovered from these output data by filtering. The CTMC will keep track of the angular movements of the vehicle and establish a launch-pad-fixed, but misaligned, reference system. This misalignment is almost constant, thus simplifying the filtering problem. Filtering is done after the PIGA outputs have been transformed by the CTMC into the strapdown reference CS. After the misalignment angles with respect to the local vertical are determined, the correction is made by rotating about the y_1- and z_1-axes (the horizontal axes).

Once the local vertical has been found and the strapdown reference CS has been leveled to the desired accuracy, the azimuth direction can be calculated from instantaneous data (measurements of the same age). Only if the azimuth measurements are too badly corrupted by noise is there a filtering problem. Because the azimuth error is the same, several calculations can be made for different times and the results can be averaged. The azimuth misalignment so determined is corrected by rotating about the x_1-axis.

The alignment is done by a rotation; therefore the direction cosines per se never have to be known. Mounting of the strapdown IMU is not critical because the alignment will correct for the inaccuracies. However, it is advantageous to start the transformation matrix with the nominal values of the direction cosines and to keep the mounting tolerances small so the alignment iterations will be as few as possible.

Inertial Sensing Instruments

The strapdown IMU utilizes three SAP's and three PIGA's (Figs. 1 and 2). The three PIGA's are oriented in the vehicle so that each, ideally, senses components of equal magnitude of the longitudinal acceleration. The accelerometer heads will rotate at about equal rates. This rotation makes certain types of errors cyclic rather than unbounded. [2,3] The SAP's are mounted to form a triad parallel to the PIGA triad for ease of PIGA output compensation of the angular velocity, optical

alignment, and packaging. [1] Under the assumption that the operation of a PIGA is known, the operation of a SAP can be described as the one for a PIGA with the pendulosity removed. The equations describing the operations of both are identical, but the dominant term in the PIGA equations becomes an error term in the SAP equations. To improve the drift characteristic, the angular momentum of the SAP is usually larger than that of the PIGA.

The instruments used for the experiment have 2^{18}-bit encoders for the SAP's and 2^{14}-bit encoders for the PIGA's. The resolution is about 5 arc sec and 0.02 m/s, respectively.

Coordinate Transformation Matrix Computer

Several characteristics of the CTMC chosen have a strong influence on the alignment procedure.

A misalignment may be corrected in two ways. One is to determine the correct set of direction cosines and replace the erroneous ones. Because the basic operation of the CTMC cannot be interrupted (or the proper updating is not done), the direction cosines must be replaced before the next updating cycle. The other way is to express the misalignment as three successive rotations about the three strapdown reference CS axes and add these during the normal operation of the CTMC.

The CTMC is of the DDA type and has a fast iteration rate (updating time is every 25.6 μs); only the second approach of misalignment correction is reasonable because the CTMC in effect aligns itself and there is no time pressure for the execution of the alignment.

The DDA type CTMC is well suited only for rectangular integration which in turn (for simplicity and therefore fast iteration time) favors the direction cosine method for updating of the transformation matrix. [4]

The word length of the CTMC is such that for alignment and other corrections the quantization level is 2^{-16} rad or 15.26 μrad (3.14 arc sec).

Updating of the Transformation Matrix Including Alignment and Orthonormalization

The transformation matrix between the strapdown instrument triad and the strapdown reference CS is defined as

$$\underline{X}_1 = C\underline{X}_2 \tag{1}$$

with

$$C \equiv \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \tag{2}$$

where $\underline{X}_1$ and $\underline{X}_2$ are the same vector, but expressed in components of the strapdown instrument CS, CS(2), and the strapdown reference CS, CS(1). The transformation matrix, ideally, is orthonormal (sometimes called orthogonal).

In general, a relative angular velocity exists between the two CS's. The angular velocity will be defined as

$$\underline{\Omega}_T \equiv \underline{\Omega} + \dot{\underline{\phi}} \tag{3}$$

where $\underline{\Omega}$ is the earth rotation rate and its components Ω_1, Ω_2, Ω_3 are known in CS(1); $\dot{\underline{\phi}}$ is the inertial angular velocity of CS(2) and its components $\dot{\phi}_1$, $\dot{\phi}_2$, $\dot{\phi}_3$ are measured by the SAP's. The rate of change of the C-matrix can then be expressed as follows [4] (see Appendix D):

$$\dot{C} = [\Omega_i]\ C - C\,[\dot{\phi}_i] \tag{4}$$

with

$$[\Omega_i] \equiv \begin{bmatrix} 0 & +\Omega_3 & -\Omega_2 \\ -\Omega_3 & 0 & +\Omega_1 \\ +\Omega_2 & -\Omega_1 & 0 \end{bmatrix}$$

and

$$[\dot{\phi}_i] \equiv \begin{bmatrix} 0 & +\dot{\phi}_3 & -\dot{\phi}_2 \\ -\dot{\phi}_3 & 0 & +\dot{\phi}_1 \\ +\dot{\phi}_2 & -\dot{\phi}_1 & 0 \end{bmatrix}$$

Equation 4 is approximated in the CTMC by the following difference equation:

$$[\Delta c_{ij}] = [\Delta_{ij}]\ C - C\,[\Delta\phi_i] \tag{5}$$

with

$$[\Delta c_{ij}] = \begin{bmatrix} \Delta c_{11} & \Delta c_{12} & \Delta c_{13} \\ \Delta c_{21} & \Delta c_{22} & \Delta c_{23} \\ \Delta c_{31} & \Delta c_{32} & \Delta c_{33} \end{bmatrix} \tag{6}$$

$$[\Delta_{ij}] = \begin{bmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} \\ \Delta_{21} & \Delta_{22} & \Delta_{23} \\ \Delta_{31} & \Delta_{32} & \Delta_{33} \end{bmatrix} \tag{7}$$

$$[\Delta\phi_i] = \begin{bmatrix} 0 & +\Delta\phi_3 & -\Delta\phi_2 \\ -\Delta\phi_3 & 0 & +\Delta\phi_1 \\ +\Delta\phi_2 & -\Delta\phi_1 & 0 \end{bmatrix} \tag{8}$$

The Δ_{ij}'s are introduced because we are now dealing with small angles rather than angular velocity; any finite

rotation, like the ones necessary for alignment and orthonormalization, can be added. Time is eliminated because the small angular changes are taken care of when they appear. We define

$$[\Delta_{ij}] = [\Omega_i]\ \Delta t - [\epsilon_{ij}] - [\delta_i] \tag{9}$$

where $[\epsilon_{ij}]$ is for orthonormalization and $[\delta_i]$ is for alignment (see Appendix A and other pertinent sections); $[\Delta\phi_i]$ and $[\Omega_i]\ \Delta t$ are of such a form that the C-matrix will always increase in magnitude. The number of times the C-matrix is operated upon could possibly give an accurate estimate of the non-normality accrued and indicate when the correction amounts to more than the lower limit set by the quantization level ($\pm 7.6\ \mu$rad or ± 1.6 arc sec).

The C-matrix is updated every time an angular increment accrues. The following formula is used for the updating:

$$C_{n+1} = C_n + [\Delta c_{ij}]. \tag{10}$$

Orthonormalization

The CTMC could conceivably be in operation for several hours before the space vehicle takes off. Since the selected transformation updating calculates each of the reference CS axes separately and independent of the other two, an accumulation of errors in orthogonality between the axes as well as in normality (unit length) of each axis is possible. Therefore a check of the orthonormality of the transformation matrix must be performed from time to time; if the necessary correction exceeds half of the angular quantization level, the correction should be made. Appendix A gives the mathematical derivation of the correction method used and the previous section showed how the correction is applied to the transformation matrix.

Determination of Leveling Misalignment and Filtering Methods Used

The local vertical as defined by a plumb bob can be determined by PIGA outputs which contain leveling misalignment data on the strapdown reference CS. Gravity causes precession of the PIGA heads; this precession rate is proportional to the gravity component along the PIGA input axis. This rate cannot be measured directly, but the head rotation angle about the input axis can be measured. An incremental encoder on the input shaft is sufficient because its output is only used to derive rate. This encoder issues a pulse every time the shaft angle is increased by the quantization level. Since the direction of the gravity vector is constant, the direction of the apparent linear velocity accumulated can be substituted to determine the gravity direction. On the pad the instruments are under the influence of wind-induced vehicle sway, vibration, and other disturbances. This causes the pulse frequency, which otherwise would be constant, to vary. Some variations can be eliminated by letting the CTMC transform the velocity pulses into the launch-pad-fixed strapdown reference CS. The other variations must be filtered out where the constancy of the direction of the local gravity vector is very beneficial.

Two approaches exist to solve the filtering problem. One has the characteristics of a closed loop and is sometimes called the implicit method. Here the intent is to eliminate the leveling error continuously; therefore, all calculations must be made in real time and will change the alignment of the transformation matrix immediately. This approach has all the characteristics of a closed loop system, e.g., damping, overshoot, etc. The other approach is open loop or explicit and relies on the fact that, in case of no-wind conditions, the misalignment is practically constant over the interval during which the alignment occurs. The misalignment is measured over a certain period and the data are evaluated later. The time needed to evaluate the misalignment is not critical. The calculated correction is then applied almost instantaneously since the strapdown IMU has a maximum angular velocity capability of 1 rad/s. The method does not have the characteristics of a closed loop; therefore, the open loop method is preferred. All data over the entire measuring period are available simultaneously for evaluation; thus filtering methods that converge much faster than those in the closed loop method are possible. This is an important point, since the alignment procedure must be compatible with the countdown procedure. Further, because the average misalignment is not really a constant but a slow function of time, a short alignment time minimizes this error influence. Various curve fits can be tried for the open loop methods; and, since the characters of the vibration and sway are predictable, the results are better.

Leveling Error Caused by Azimuth Error

An error in azimuth causes an error in leveling. The error is caused by inaccurate precessing of the strapdown reference CS. The precessing is done to keep the reference CS aligned with CS(0) which is rotating with the earth.

The error has the following form (see Appendix C for its derivation):

$$\delta_{z7A} = \cos\phi_0 \Omega t \delta_A \tag{11}$$

where CS(7) is a CS with the x-axis parallel to the local vertical and the z-axis pointing due east. The angle Ωt is assumed to be small; ϕ_0 is the geodetic latitude, and δ_A is the azimuth error.

Numerical evaluation of the error in leveling for the Saturn V Merritt Island launch sites results in

$$\delta_{z7A} = 64.1\ 10^{-6} t\delta_A \text{ rad.}$$

The final overall leveling error should not be more than 100 μrad (20 arc sec). About 5 μrad can be allocated to the leveling error caused by an azimuth error. Thus for the final leveling (during the last iteration), the error in azimuth should be below 1.3 mrad or 4.5 arc min if an alignment time of 60 seconds is assumed. At least two alignment iterations are anticipated.

Determination of Azimuth Misalignment

The final azimuth alignment should only be done after the strapdown reference CS is leveled because of the strong influence of the leveling error on the azimuth error (discussed in the next section). Under these conditions the line of sight (LOS) from the strapdown IMU or a unit vector along it can be expressed in both the launch pad coordinate system, CS(0), and the strapdown reference CS, CS(1). The depression γ_0 and the azimuth α_0 of the LOS are accurately known in CS(0). The components should be the same in CS(1) and the difference can be used to calculate the necessary azimuth correction. This calculation can be done at leisure as long as data of the same age are used. For confidence or elimination of noise the calculation can be repeated at different times; and, since the error should be the same, the results can be averaged. The LOS is established by the theodolite on the ground and a porro prism on the strapdown IMU. The prism is servoed about the strapdown IMU x_4-axis.

The equations governing the azimuth error determination are outlined as follows. The unit vector of the LOS is expressed in CS(6) and then transformed into CS(1):

$$\begin{bmatrix} L_{x1} \\ L_{y1} \\ L_{z1} \end{bmatrix} = {}^1C_6 \begin{bmatrix} L_{x6} \\ L_{y6} \\ L_{z6} \end{bmatrix} \qquad (12)$$

where

$${}^1C_6 = {}^1C_2\,{}^2C_3\,{}^3C_4\,{}^4C_5\,{}^5C_6 .$$

The nomenclature is defined in Appendix B. For any one azimuth error determination, the transformation ${}^1C_6$ is constant; therefore,

$${}^1C_6 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} . \qquad (13)$$

Substituting equation 13 and angle γ_p into equation 12 yields

$$\begin{bmatrix} L_{x1} \\ L_{y1} \\ L_{z1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} -\sin\gamma_p \\ +\cos\gamma_p \\ 0 \end{bmatrix} \qquad (14)$$

where γ_p is the depression of the LOS expressed in CS(6). This angle is not known, but can be determined by the following considerations. Since CS(1) is already leveled, L_{x1} and L_{x0} are identical, leading to

$$L_{x1} = L_{x0} = -\sin\gamma_0$$

or

$$\sin\gamma_0 = a_{11}\sin\gamma_p - a_{12}\cos\gamma_p . \qquad (15)$$

This expression can be solved for $\sin\gamma_p$ and $\cos\gamma_p$ without the determination of the angle itself. With everything known in equation 14, we can determine L_{y1} and L_{z1} from which the misaligned azimuth angle of the LOS in CS(1) is obtained by

$$\alpha_1 = \frac{\pi}{2} + \tan^{-1}\frac{L_{y1}}{L_{z1}} . \qquad (16)$$

An azimuth correction must be made of

$$\delta_A = \alpha_1 - \alpha_0$$

where α_0 is the azimuth angle in CS(0).

If the azimuth α_0 changes as a function of time, it can be broken into several parts

$$\alpha_0 = \alpha_R + \alpha(t), \qquad (17)$$

i.e., into a reference azimuth and a time varying portion. The alignment can be made with respect to the reference azimuth and the time varying portion can be added later without going through the full azimuth alignment determination. This affords a very convenient way of implementing a continuously varying launch azimuth.

The azimuth determination for the earlier alignment iterations can use the same procedure as outlined previously for the final iteration. This procedure will not result in an accurate azimuth for the early iterations, but it will improve the azimuth estimate and be sufficient for the early iterations.

Azimuth Error Caused by Leveling Error

An error in leveling will cause an error in the azimuth determination because the LOS from the strapdown IMU to the azimuth theodolite has a sizable depression (about 27 degrees for Saturn V). The relationship is

$$\delta_A = \tan\gamma_0\, \delta_L = 0.51\, \delta_L$$

where γ_0 is the LOS depression and δ_L is the leveling error about the projection of the LOS into the horizontal plane. The azimuth error is not affected by an error in leveling perpendicular to the LOS. This high sensitivity indicates that the final azimuth alignment can only be done after the final leveling. The fact that the LOS as seen from the vehicle has a built-in leveling error is taken care of in the equations and does not contribute to the above relationship.

Conclusions

The strapdown reference system (IMU and CTMC) can be used to advantage for the initial alignment where the strapdown system calculates a misaligned, but launch-pad-fixed, reference CS. This active utilization per se constitutes a prelaunch check for proper operation of the strapdown system. Relatively large misalignments of the parameters connected with the porro prism (used for azimuth determination) can be tolerated as long as they are constant and well known (optically determined). This relaxation of the manufacturing tolerances was made possible by the use of compensation matrices during the determination of the azimuth error. The only critical alignment concerns the mutual orthogonality of the inertial sensing instrument axes and the parallelity of the PIGA and the SAP triad.

Appendix A. Mathematical Derivation of the Orthonormalization Method

Vector components are transformed from the strapdown inertial sensing instrument CS to the strapdown reference CS by the C-matrix, which is updated by the CTMC with

$$[\Delta c_{ij}] = [\Delta_{ij}]\ C - C[\Delta\phi_i]$$

where

$$[\Delta c_{ij}] = \begin{bmatrix} \Delta c_{11} & \Delta c_{12} & \Delta c_{13} \\ \Delta c_{21} & \Delta c_{22} & \Delta c_{23} \\ \Delta c_{31} & \Delta c_{32} & \Delta c_{33} \end{bmatrix}$$

$$[\Delta_{ij}] = \begin{bmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} \\ \Delta_{21} & \Delta_{22} & \Delta_{23} \\ \Delta_{31} & \Delta_{32} & \Delta_{33} \end{bmatrix}$$

$$[\Delta\phi_i] = \begin{bmatrix} 0 & +\Delta\phi_3 & -\Delta\phi_2 \\ -\Delta\phi_3 & 0 & +\Delta\phi_1 \\ +\Delta\phi_2 & -\Delta\phi_1 & 0 \end{bmatrix}$$

A detailed discussion of this formula was previously given. These difference equations non-orthonormalize the matrix after many iterations. An orthonormalization method must be applied to eliminate the accumulated error whenever it exceeds half of the quantization level.

To develop a correction method, the orthonormality condition

$$A^T A = I \tag{A1}$$

is used, where A is a perfectly orthonormal matrix (normally only called orthogonal matrix with the normality implied) and I is the three-by-three identity matrix. We assume

$$A = (I - \epsilon)C \tag{A2}$$

where ϵ is the correction matrix which transforms the non-orthonormal matrix C into an orthonormal matrix A. The following discussion could as well be developed for $A = C(I - \epsilon)$ without changing the basic approach.

Substitution of equation A2 into equation A1 results in

$$C^T(I - \epsilon^T)\ (I - \epsilon)C = I$$

or

$$(I - \epsilon^T - \epsilon + \epsilon^T\epsilon) = (C^{-1})^T C^{-1}. \tag{A3}$$

Relationship A2 does not uniquely define A which can be shown as follows: Assume that one solution is known.

$$A_0 = (I - \epsilon_0)C. \tag{A2a}$$

We can transform A_0 into any one of all possible orthonormal matrices by a rotation matrix R

$$A = RA_0 = R(I - \epsilon_0)C, \tag{A4}$$

but A satisfies condition A2, i.e.,

$$A = (I - \epsilon)C.$$

This shows that a multitude of solutions exist for ϵ and an additional condition must be found. Among all possible orthonormal matrices, one describes a "best fit" to the non-orthonormal matrix. As a criterion for the "best fit" the minimum of the norm of the matrix ϵ has been selected, the norm being

$$||\epsilon|| = +\left\{ \sum_{i=1,j=1}^{3} \sum^{3} \epsilon_{ij}^{\,2} \right\}^{1/2} \tag{A5}$$

Sustituting equation A2 into equation A4 results in

$$(I - \epsilon) = R(I - \epsilon_0)$$

or

$$\epsilon = R\epsilon_0 - (R - I). \tag{A6}$$

To evaluate the norm we must work with the matrix elements; ϵ_0 is defined as

$$\epsilon_0 \equiv \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix} \tag{A7}$$

This definition is not in conflict with equation A5 because when the norm is evaluated the rotation matrix is set to zero.

To investigate the behavior of ϵ in the vicinity of ϵ_0, we assume that R is an infinitesimal rotation consisting of three infinitesimal mutually orthogonal rotations δ_i (i = 1, 2, 3) and set

$$R = \begin{bmatrix} 1 & +\delta_3 & -\delta_2 \\ -\delta_3 & 1 & +\delta_1 \\ +\delta_2 & -\delta_1 & 1 \end{bmatrix} \tag{A8}$$

When the norm is a minimum, the square of the norm is also a minimum; evaluating the square of the norm is simpler and results in the following expressions for the minimum:

$$\frac{\partial}{\partial\delta_1} ||\epsilon||^2 = 2\epsilon_{32} - 2\epsilon_{23}$$

$$\frac{\partial}{\partial\delta_2} ||\epsilon||^2 = 2\epsilon_{13} - 2\epsilon_{31}$$

$$\frac{\partial}{\partial\delta_3} ||\epsilon||^2 = 2\epsilon_{21} - 2\epsilon_{12}$$

for

$$\delta_1 = \delta_2 = \delta_3 = 0.$$

This result indicates that

$$\epsilon = \epsilon^T \tag{A9}$$

for the norm of ϵ to be a minimum.

Substituting equation A9 into equation A3 and solving for ϵ yields

$$\epsilon = I - \sqrt{(C^{-1})^T C^{-1}}. \tag{A10}$$

For small ϵ equation A10 can be solved iteratively by

$$\epsilon_n = 0.5[I - (C^{-1})^T C^{-1}] + 0.5\epsilon_{n-1}^2. \tag{A11}$$

The iteration is stopped when ϵ_n and ϵ_{n-1} differ by less than a specified quantity. The orthonormal matrix is then

$$A = (I - \epsilon_n)C. \tag{A2b}$$

Another orthonormalization method, equivalent to the one previously discussed, is iterative with the following definitions:

$$\epsilon_n' \equiv 0.5(C_n C_n^T - I) \tag{A12}$$

with

$$\epsilon_1' \equiv 0.5(CC^T - I) \tag{A13}$$

and

$$C_n = (I - \epsilon_{n-1}')C_{n-1}. \tag{A14}$$

Infinite iterations result in

$$A = C_\infty = \left[\prod_{n=1}^{\infty} (I - \epsilon_n')\right] C. \tag{A15}$$

If ϵ_1' is small, very few (probably only one) iterations are necessary for the desired accuracy, and the problem of matrix inversion of the earlier method is eliminated.

The equivalence of equations A2, with the condition that $\epsilon = \epsilon^T$, and A15 can be demonstrated from the definitions of the correction matrices. Squaring and regrouping of equation A10 results in

$$CC^T = [(I - \epsilon)^2]^{-1}.$$

This expression can be expanded into a series

$$CC^T = I + 2\epsilon + 3\epsilon^2 + 4\epsilon^3 + 5\epsilon^4 + \ldots.$$

or, with equation A13,

$$\epsilon_1' = 0.5(CC^T - I) = \epsilon + \frac{3}{2}\epsilon^2 + 2\epsilon^3 + \frac{5}{2}\epsilon^4 + 3\epsilon^5 + \ldots \tag{A16}$$

Inversion of the series leads to

$$\epsilon = \epsilon_1' - \frac{3}{2}\epsilon_1'^2 + \frac{5}{2}\epsilon_1'^3 - \frac{35}{8}\epsilon_1'^4 + \frac{63}{8}\epsilon_1'^5 - \frac{231}{16}\epsilon_1'^6 + \frac{429}{16}\epsilon_1'^7 \ldots \tag{A17}$$

Remaining to be demonstrated is

$$I - \epsilon = \prod_{n=1}^{\infty} (I - \epsilon_n').$$

First, ϵ_n' is expressed as a function of ϵ_{n-1}':

$$\epsilon_n' = 0.5(C_n C_n^T - I)$$

$$= 0.5[(I - \epsilon_{n-1}')C_{n-1}C_{n-1}^T(I - \epsilon_{n-1}') - I]$$

$$= 0.5[(I - \epsilon_{n-1}')(I + \epsilon_{n-1}')(I - \epsilon_{n-1}) - I]$$

$$\epsilon_n' = \epsilon_{n-1}'^3 - \frac{3}{2}\epsilon_{n-1}'^2 .$$

With this recursion formula we develop

$$\prod_{n=1}^{\infty}(I - \epsilon_n') = I - \epsilon_1' + \frac{3}{2}\epsilon_1'^2 - \frac{5}{2}\epsilon_1'^3 + \frac{35}{8}\epsilon_1'^4 - \ldots$$

The right-hand side is equivalent to $(I - \epsilon)$ (equation A17), and the equivalence of the two orthonormalization methods has been proven.

Appendix B. Coordinate Transformations

Coordinate transformations are indicated by ${}^bC_a$ where the transformation is performed from CS(a) to CS(b) or

$$\begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} = {}^bC_a \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix} .$$

In particular the following definitions and comments apply.

$${}^1C_2 \equiv \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \equiv C \equiv [c_{ij}] .$$

This transformation matrix is updated by the CTMC using angular rates about vehicle axes, earth rotation rate compensation, etc.

$${}^2C_3 = \begin{bmatrix} +\sqrt{1/3} & +\sqrt{1/2} & +\sqrt{1/6} \\ +\sqrt{1/3} & -\sqrt{1/2} & +\sqrt{1/6} \\ +\sqrt{1/3} & 0 & -2\sqrt{1/6} \end{bmatrix} .$$

This transformation matrix is ideal (any error is reflected into ${}^1C_2$), and the values are chosen so that the PIGA's sense components of equal magnitude of the longitudinal acceleration.

$${}^3C_4 = \begin{bmatrix} 1 & -\delta_{z3} & +\delta_{y3} \\ +\delta_{z3} & 1 & 0 \\ -\delta_{y3} & 0 & 1 \end{bmatrix} .$$

The δ_i's are small misalignments of the porro prism rotation axis.

$${}^4C_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_p + \delta_{x4}) & -\sin(\alpha_p + \delta_{x4}) \\ 0 & \sin(\alpha_p + \delta_{x4}) & \cos(\alpha_p + \delta_{x4}) \end{bmatrix} ,$$

and

$${}^5C_6 = \begin{bmatrix} 1 & 0 & +\delta_{y5} \\ 0 & 1 & 0 \\ -\delta_{y5} & 0 & 1 \end{bmatrix} .$$

See Nomenclature for explanation of the angles.

All error angles are identified by the symbol δ, which may have up to three subscripts, i, j, and k, and in this order if more than one is used. The subscript i can take on the letters x, y, and z identifying the axis about which the error angle occurs. The subscript j identifies the pertinent CS and will be a number between 0 and 7. The subscript k identifies the source of the error angle and will be named by a letter and explained as it occurs. The polarity of the angular error is defined as

$$(\text{actual angle}) = (\text{ideal angle}) + \delta_{ijk} .$$

A positive error angle is a rotation in the mathematically positive sense away from the ideal location.

Appendix C. Derivation of the Leveling Error Caused by Error in Azimuth

To remain launch-pad-fixed, the strapdown inertial reference CS must be rotated with the earth rotation rate, which is split into components of angular velocity about the strapdown reference CS, CS(1). These components will be in error when the azimuth of the strapdown reference CS is in error. For the following discussion, assume that the strapdown reference CS is perfectly level at the start of the evaluation of the precession error (it will be shown that the leveling error is a function of time). The error in leveling does not have an influence on the azimuth error if there is no

updating of the azimuth estimate during the time of the leveling. The effect of the error in azimuth is such that the strapdown reference CS is rotated about an axis slightly misaligned with the earth rotation axis. The magnitude of the earth rotation rate is well known and not assumed in error. For a rotation about an axis with constant direction cosines, we have the following relationship for the transformation matrix at time t, C_t, with C_0 being the transformation matrix at t = 0 (Appendix D):

$$C_t = \Phi C_0 \qquad (C1)$$

where

$$\Phi \equiv \begin{bmatrix} d_1^2 v\phi + c\phi & d_1 d_2 v\phi + d_3 s\phi & d_1 d_3 v\phi - d_2 s\phi \\ d_2 d_1 v\phi - d_3 s\phi & d_2^2 v\phi + c\phi & d_2 d_3 v\phi + d_1 s\phi \\ d_3 d_1 v\phi + d_2 s\phi & d_3 d_2 v\phi - d_1 s\phi & d_3^2 v\phi + c\phi \end{bmatrix}$$

with $\phi = \phi(t)$; $s\phi \equiv \sin\phi$; $v\phi \equiv \text{vers}\,\phi = 1 - \cos\phi$; $c\phi \equiv \cos\phi$. The d's are the direction cosines of the rotation axis with respect to either CS between which C_t transforms. Equation C1 holds for any C_0 and the identity matrix is selected for convenience. As reference we select a CS inertially-fixed and coinciding with both CS(0) and CS(1) at t = 0. The transformation between CS(0) and CS(1) takes the form:

$$^1C_0 = {}^1C_R \, {}^RC_0$$

with

$$^1C_R = \Phi_1 \text{ and } {}^RC_0 = (\Phi_0)^T .$$

The direction cosine between the x-axes in both CS(1) and CS(0) will indicate the leveling error of CS(1) caused by the erroneous precession. For evaluation, we work with the matrix elements to obtain:

$$\text{For } \Phi_p: \quad d_{10} = +\sin\phi_0, \quad d_{20} = -\cos\phi_0, \quad d_{30} = 0$$

$$\text{For } \Phi_1: \quad d_{11} = +\sin\phi_0, \quad d_{21} = -\cos\phi_0\cos\delta_A, \quad d_{31} = +\cos\phi_0\sin\delta_A$$

where ϕ_0 is the geodetic latitude of the launch pad and δ_A is the azimuth error. The rotation is time dependent: $\phi = \Omega t$, where Ω is the earth rotation rate.

Evaluation of the first element of the first column of $^1C_0$ leads to

$$a_{11} = (s^2\phi_0 v\phi + c\phi)^2 + (s\phi_0 c\phi_0 c\delta_A v\phi - c\phi_0 s\delta_A s\phi) s\phi_0 c\phi_0 v\phi + (s\phi_0 c\phi_0 s\delta_A v\phi + c\phi_0 c\delta_A s\phi) c\phi_0 s\phi .$$

This element is equivalent to the cosine of the leveling error

$$\cos\delta_{zA} = a_{11} = 1 - 0.5\delta_{zA}^2 .$$

The latter approximation can be made since the error is small. Evaluation leads to

$$\delta_{zA}^2 = \delta_A^2 c^2\phi_0 (s^2\phi_0 v^2\phi + s^2\phi) .$$

Since $\phi = \Omega t$ is small,

$$\delta_{zA} = \cos\phi_0 \Omega t \delta_A .$$

This error is a function of time and, therefore, the strapdown reference CS misalignment is also a function of time. The assumption of a linear function of time is very good and will be used in the evaluation of the misalignment terms. The leveling error caused by the erroneous precession is (neglecting second order terms) in the meridian plane of the launch site.

Appendix D. Coordinate System Rotation About an Arbitrary, but Fixed, Axis

Euler's theorem states that any physically possible rotation of a body from one orientation to another can be expressed as a single rotation through the angle ϕ about a fixed axis with the direction cosines d_1, d_2, and d_3. The d's are the same for both coordinate systems and remain constant. The angle ϕ does not have any restrictions and can be a function of time. If the body rotation is expressed as a CS transformation,

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \Phi \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

where

$$\Phi = \begin{bmatrix} d_1^2 v\phi + c\phi & d_1 d_2 v\phi + d_3 s\phi & d_1 d_3 v\phi - d_2 s\phi \\ d_2 d_1 v\phi - d_3 s\phi & d_2^2 v\phi + c\phi & d_2 d_3 v\phi + d_1 s\phi \\ d_3 d_1 v\phi + d_2 s\phi & d_3 d_2 v\phi - d_1 s\phi & d_3^2 v\phi + c\phi \end{bmatrix} \qquad (D1)$$

with $v\phi \equiv \text{vers}\,\phi = 1 - \cos\phi$; $c\phi \equiv \cos\phi$; $s\phi \equiv \sin\phi$; and $\phi = \phi(t)$.

The updating law for the strapdown system transformation matrix can be extracted from equation D1 by calculating the rate of change of Φ when ϕ is zero:

$$d\Phi/dt = \begin{bmatrix} 0 & +\dot{\phi}_3 & -\dot{\phi}_2 \\ -\dot{\phi}_3 & 0 & +\dot{\phi}_1 \\ +\dot{\phi}_2 & -\dot{\phi}_1 & 0 \end{bmatrix} = [\dot{\phi}_i]$$

with $\dot{\phi}_i = d_i\dot{\phi}$. The $\dot{\phi}_i$'s are the components of the angular velocity between the CS's.

References

1. Daniel, F. P.; Doane, G. B., III; and Kissel, R. R.: Application of PIGA's and SAP's to Strapdown Inertial Navigation at MSFC, Presented at the AIAA Guidance, Control, and Flight Dynamics Conference, Aug. 14-16, 1967, Huntsville, Alabama (Conf.). Published as NASA TM X-53619, July 7, 1967.

2. Otten, D. D.: Body-Fixed, Three-Axis Reference System Study, Phase I Final Report, Thompson Ramo Wooldridge Systems Rpt. No. 4499-6007-R0000, May 2, 1966.

3. Body-Fixed, Three-Axis Reference System Study, Phase II Final Report, Thompson Ramo Wooldridge Systems Rpt. No. 05128-6001-R000, Dec. 15, 1966.

4. Burdeshaw, Dexter H.: Methods of Computing the Transformation Matrix Associated with Gimballess Inertial Measurement Units, NASA TM X-53294, July 13, 1965.

FIGURE 1. MOCKUP OF STRAPDOWN IMU (TOP VIEW).

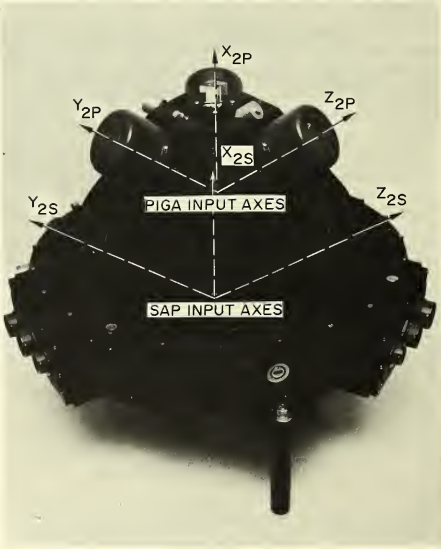


FIGURE 2. MOCKUP OF STRAPDOWN IMU (SIDE VIEW).

No. 67-569

AUTOMATIC FLIGHT CONTROL SYSTEM FOR AUTOMATIC TERRAIN FOLLOWING

by

F. M. KRACHMALNICK, G. J. VETSCH,
and
M. J. WENDL
McDonnell Douglas Corporation
St. Louis, Missouri

AIAA Paper
No. 67-569

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

2.09, 8.07, 8.08

-- NOTES --

AUTOMATIC FLIGHT CONTROL SYSTEM FOR AUTOMATIC TERRAIN FOLLOWING*

F. M. Krachmalnick, Section Manager
G. J. Vetsch, Senior Engineer
M. J. Wendl, Senior Engineer

Guidance and Control Mechanics Department
McDonnell Douglas Corporation
St. Louis, Missouri

Abstract

The background and goals of the Project 666A Automatic Terrain Following Program are described and the responsibilities assigned to the associate contractors are presented. The functional operation and features of the high performance, highly reliable fixed gain Automatic Flight Control System (AFCS) are described. Results from digital and analog longitudinal three degree-of-freedom simulation system studies are compared with the currently proposed NADC and C* handling qualities criteria. A description of the six degree-of-freedom nonlinear simulation of the closed loop airframe/AFCS/Forward Looking Radar combination is summarized. Terrain following performance and radar system failure effects study results are presented. The results obtained from the AFCS flight test development program performed at McDonnell are discussed and comparisons of flight test results with simulation results are provided.

I. Introduction

1.0 Background and Goals of 666A Program

McDonnell was selected as the airframe contractor in Project 666A, a portion of ADO-29, funded by the Aeronautical Systems Division and supported by the Systems Technology Division at Wright-Patterson Air Force Base (WPAFB). This program was conducted at McDonnell in the time period of June 1965 through December 1966. The USAF goals of Project 666A were to develop and demonstrate an automatic terrain following capability in the vertical plane, and to provide precise lateral control for guidance and navigational course direction in the horizontal plane of a high performance, fighter-bomber type of aircraft, such as the McDonnell F-4.

2.0 Performance and Reliability Requirements for AFCS in LAHS Environment

The Air Force desires to develop an automatic terrain following capability for penetrating modern anti-aircraft defense systems which are extremely effective at all portions of the flight envelope with the exception of high speed, low altitude flight (LAHS). The probability of aircraft survival and mission completion is increased if the penetrating aircraft can remain undetected by these enemy defenses. Radar detection can be avoided by the aircraft flying fast and low using the shielding provided by the natural terrain. The design of a terrain following command system must provide this terrain following capability with essentially zero probability of impact with the ground.

McDonnell's flight experience with manual terrain following in the RF-4C aircraft has shown that a pilot cannot fly safely and perform an accurate task of terrain following at very low altitudes for extended periods of time due to fatigue and cockpit workload. The need for an automatic terrain following capability is clearly indicated. In order to accomplish the automatic task, the requirement for a high performance, highly reliable automatic flight control system is mandatory to insure a high probability of penetration and mission success.

3.0 Responsibilities of Associate Contractors

McDonnell, General Electric, and Texas Instruments were selected by the USAF to participate in the automatic terrain following effort. Texas Instruments was responsible for the design, development, and fabrication of the AN/APQ-101 Forward Looking Radar and the Terrain Following Computer. General Electric was given responsibility for the design and fabrication of the AFCS hardware, performing analytical design studies, accomplishing flight-worthiness tests of the AFCS hardware, and for participating in the flight test evaluation of the AFCS.

McDonnell's responsibilities in the 666A Program were to:

(a) participate in the design and development of the 666A Automatic Flight Control System (AFCS);

(b) make provisions for the wiring and installation of the General Electric AFCS and the Texas Instruments' Forward Looking Radar (AN/APQ-101) into an F-4 Aircraft;

(c) perform flight-worthiness tests on the G.E. redundant integrated servo actuator using the McDonnell "iron bird" test facilities;

(d) install and integrate the AFCS and radar into the aircraft;

(e) perform the flight test evaluation of the AFCS.

This paper summarizes the work accomplished by McDonnell in accordance with its assigned tasks in the program. Special emphasis is placed on the analysis, simulation, and flight test efforts expended in developing and evaluating the AFCS for use in providing the automatic terrain following capability. Simulations and flight tests were conducted to determine whether the performance of the AFCS will meet the specified design requirements.

II. 666A Automatic Flight Control System (AFCS)

The 666A AFCS was designed to provide the high degree of performance and reliability required for automatic terrain following in the sensitive low altitude high speed (LAHS) environment.

1.0 Features of the 666A AFCS

The 666A AFCS provides the following features in pitch, roll, and yaw channels.

* Work funded by Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, under Contract AF33(657)-13543.

1.1 Pitch Channel. Pitch channel features are:

(a) high performance throughout the F-4 flight envelope using fixed gains in all control modes of operation. Conventional autopilots in current service aircraft utilize air data computers to control the gain changing functions. In more recent years, the concept of adaptive gain changing has appeared more promising for providing improved and uniform handling qualities. Use of either type of gain changing technique has resulted in a reduction in reliability and an increase in maintenance time for the automatic flight control system. The 666A AFCS uses fixed gains in all control modes of operation, and, therefore, the reliability and ease of maintainability of the AFCS should be vastly increased.

(b) nearly invariant stick force per "g" and nearly uniform dynamic response handling qualities characteristics throughout the entire flight envelope.

(c) three channels of identical sensors and electronics with voting to provide failure detection and correction for the electronics and the three main ram feedback signals from the stabilator actuator.

(d) fail-operate capability in the electronics with a calculated MTBF in excess of 50,000 hours based on a 5 hour mission that includes one hour of automatic terrain following.

(e) fail-operate, fail-neutral operation of the stabilator actuator. Dual servo actuators with electrical feedback and an electrical "model" provide failure detection and correction.

(f) control augmentation and normal acceleration command inputs.

(g) blended pitch rate and normal acceleration feedback.

(h) series or parallel servo actuation with automatic trim available in both servo modes. Both servo configurations were incorporated into the AFCS to provide a test article for evaluating which configuration would provide the better performance relative to handling qualities and terrain following. In the parallel mode, the servo moves the control stick and the entire control system. The parallel servo deflection authority allowed is 100%, i.e., the full 30 degrees of stabilator surface motion. However, the servo is torque-limited and can be overpowered with an aft stick force of 11 pounds and a push force of 14 pounds. In the series mode, the actual servo deflection does not move the control stick. The series servo is deflection-limited to $\pm$ 2 degrees of equivalent stabilator motion. The servo can be overpowered when the pilot moves the control stick in a direction opposite to the servo deflections. Both servo configurations can drive the stabilator at rates up to 30 deg/sec within their deflection authority limits.

Automatic trim is used to unload the servo in the parallel mode and to effectively increase the deflection authority in the series mode. Auto-trim maintains near zero steady state forces for the parallel servo and is energized whenever the linkage force between the parallel servo and the manual control system exceeds $\pm$ 2.5 pounds. The trim motor runs in a direction to force balance the manual control system and thereby relieves the parallel servo loads. Auto-trim in the series servo mode is used to extend the authority of the deflection-limited series servo. Trim is initiated whenever the servo ram displacement exceeds a set threshold relative to its centered position. The trim motor runs to force unbalance the manual control system causing the control surface to move to the commanded position and results in the servo being recentered in the steady state condition. The automatic trim motor rate in the F-4 is approximately one deg per second of equivalent stabilator rate and is commensurate with safety-of-flight criteria established for the aircraft.

1.2 Roll and Yaw Channels. The lateral channels are virtually unchanged from the production F-4 AFCS except with respect to safety considerations. Roll channel features are:

(a) dual sensors, electronics and series actuators with a monitor used to provide failure detection and shut-down.

(b) reduced servo authority from $\pm$ 7.5° to $\pm$ 3° aileron and fast-centering time (changed from present 3 to 10 seconds to 0.3 seconds).

(c) fixed gains in all modes of operation.

Yaw channel features are:

(a) dual sensors and electronics to provide an active channel and "model" channel to enable failure detection and shut-down.

(b) fixed gains in all modes of operation.

2.0 666A Program AFCS Modes of Operation

The modes of AFCS operation are:

- o Stability (Control) Augmentation:
  Pitch, roll, and yaw short period damping and lateral turn-coordination.
- o Pilot Assist (with Control Stick Steering):
  Pitch: Attitude hold, Altitude hold.
  Roll: Attitude hold, Ground track hold.
- o Automatic Terrain Following:
  Automatic pitch path control using either manual stability-augmented lateral control or automatic ground track hold as may be selected by the pilot.

3.0 AFCS Hardware

The 666A AFCS hardware includes the following components:

(a) Pitch Computer
(b) Roll-Yaw Computer
(c) Redundant hydraulic stabilator actuator
(d) Fast-centering roll servo actuators
(e) Stick force transducer
(f) Engaging controller
(g) Monitor panel
(h) Sensors

The test equipment for ground checkout and troubleshooting consists of a pre-flight test panel and associated power supplies.

4.0 Functional Description of AFCS

The requirement stipulated by the USAF for the AFCS design was that no single malfunction, whether in the automatic flight control or in other systems, such as the hydraulic and electrical supplies, will generate a situation that requires immediate and/or precise action by the pilot to avoid disaster. This requirement was applicable to all modes of operation and all channels of control.

4.1 Pitch Channel. The function of the pitch channel of the AFCS is to provide acceptable flight path response in the automatic terrain following mode and satisfactory handling qualities in the stability (control) augmentation and pilot assist modes. A functional block diagram of the pitch axis is presented in Figure 1.

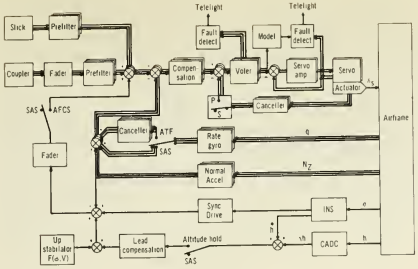


Figure 1 - AFCS Pitch Axis Block Diagram

All sensors and electronics of the inner loop, as well as the signals from the stick force transducer and the terrain following coupler, are provided in triplicate to increase the reliability of this channel. The voter on the electronics and the three main ram feedback signals insures that the outputs applied to the servo amplifiers and to the model are good signals despite any one failure occurring in any of the three channels of the electronics preceding the voter. A failure is detected and indicated to the pilot in the form of a light on the telelight panel. A second voter is utilized to detect any failure on the outputs of the auxiliary rams of the servos. Deviations above a set tolerance among the two auxiliary ram position signals and the model electrical output are detected, the failed channel is shut down, and the failure is indicated to the pilot. A second failure in either the model or a remaining servo ram channel causes a center lock mechanism to cam the servo ram to its null position. This centering action prevents a servo hardover in the event of a second failure. Series or parallel servo operation is available for pilot selection in all modes.

Blended pitch rate and normal acceleration are used as feedbacks to all input commands with the exception of automatic terrain following where pitch rate feedback is cancelled to accommodate the ride control and g-limits of the automatic terrain following (ATF) command computer. The attitude command signal is derived from the Inertial Navigation System (INS) and the altitude command signal from the Central Air Data Computer (CADC). An up-stabilator signal as a function of bank angle and velocity is commanded in these modes to preclude loss in attitude and altitude during turns.

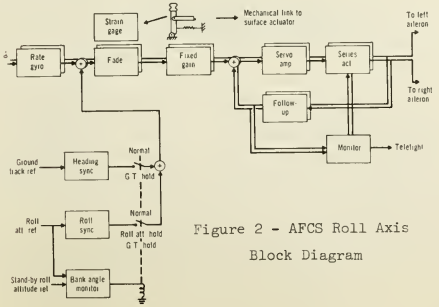


Figure 2 - AFCS Roll Axis Block Diagram

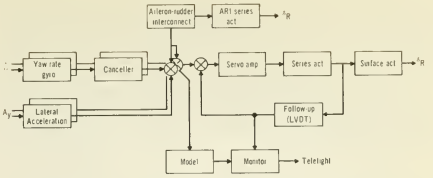


Figure 3 - AFCS Yaw Axis Block Diagram

4.2 Roll Channel. The function of the roll channel of the AFCS is to reduce roll-yaw coupling by utilizing roll damping to alleviate disturbance-induced roll rates while providing good roll response to pilot commands. The functional block diagram of this mode is provided in Figure 2. Roll attitude hold and automatic ground track hold are provided as the pilot assist modes. The servo ram position outputs are continuously compared and if the difference exceeds 50% of full servo stroke, the prescribed tolerance, a fail light is illuminated in the telelight panel, and the roll stability augmentation mode is disengaged. In the terrain following mode, either the stability augmentation or automatic ground track hold mode may be selected. The automatic ground track mode incorporates a bank angle monitor, which compares the INS roll attitude signal with the roll attitude signal derived from a separate miniature vertical gyro source, used with the forward looking radar, to provide an increase in reliability for the roll signal. If these two signals differ by more than 5 degrees of bank angle, the automatic ground track hold mode will not engage. This mode will also not engage if either roll attitude source indicates a bank angle exceeding 15 degrees.

4.3 Yaw Channel. The function of the yaw channel is to improve dutch-roll damping and directional stability of the basic aircraft. The functional block diagram of this channel is presented in Figure 3. Monitoring consists of comparing the yaw servo position with the electrical output of an electronic model representative of the servo loop dynamics with the shut-down technique being identical to that used in the roll channel. Incorporation of the lateral accelerometer provides automatic turn-coordination in this mode. Aileron-to-rudder interconnect is incorporated to preclude adverse yaw due to aileron deflections at low speed flight conditions.

III. AFCS Analysis and Simulation Results

1.0 Analysis

Since the lateral and directional channels of the 666A are functionally similar to those of the current F-4 (FC8) AFCS, major emphasis was placed on the analysis of the longitudinal channel.

The design philosophy of the 666A AFCS is based on a modified sensitivity concept whereby the necessary control system compensation is systematically chosen to minimize system sensitivity to plant variations and to random disturbance inputs. The synthesis technique used in the design of the fixed-gain system results in network compensation which, when combined with the aircraft (plant) open loop frequency response characteristics, insures stability margins above a prescribed minimum throughout the flight regime. The manner in which the system

constraints and plant variations are assimilated and used in obtaining an optimum compensation loop definition is unique to this technique. The design process allows efficient use of the digital computer facilities for processing the routine calculations involving airframe dynamics and influential feedback parameters, such as the canceller time constant and the blended normal acceleration to pitch rate feedback gain ratio.

In addition to meeting stability requirements, the pitch channel was designed to essentially meet the performance criterion proposed by the Naval Air Development Center (NADC) in NADC-ED-6282, and the stick force per "g" gradient required by MIL-F-8785. Since the primary use of the AFCS was in the LAHS flight regime for accomplishing the automatic terrain following task, the fixed-gain longitudinal system design was directed such that the handling qualities would meet the NADC criterion for the LAHS flight conditions, which include the flight regime from .8 Mach to V_{max} for altitudes below 15,000 feet. Special emphasis was placed on tailoring the system to meet this criterion as closely as possible for the higher altitude and low "q" flight conditions. The NADC criterion is compared with the MIL-F-8785 and the Boeing C* criterion in Figure 4 in an equivalent second order frequency and damping ratio format. For evaluating the 666A AFCS, the NADC criteria was converted to a normalized time history envelope of performance acceptability and is presented along with the C* time history envelopes in Figure 5. Such envelopes enable meaningful comparisons of augmented aircraft responses, which may be slightly nonlinear or of higher order, to be made with the criteria.

The NADC proposed revision to MIL-F-8785 specifies that the longitudinal short period dynamic oscillations of normal acceleration, in equivalent damping and frequency, be used for comparison with the criterion requirements. McDonnell has extrapolated the NADC criterion to reflect the measure of pilot's response to the blend of pitch rate and normal acceleration as sensed at the pilot's seat location. For aircraft velocities above 400 feet/second, the responses used for criterion comparison should be normal acceleration measured at the pilot's seat; for velocities below 400 feet/second, pitch rate responses should be used for criterion comparison; and for velocities in the vicinity of 400 feet/second, both pitch rate and normal acceleration at the pilot location (N_{z_p}) responses should fall within the criterion envelope.

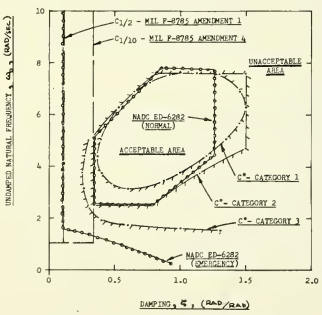


Figure 4 - Longitudinal Handling Qualities Criteria

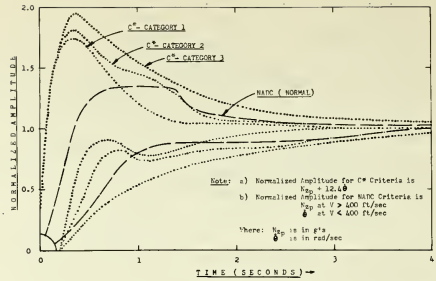


Figure 5 - Normalized Time History Response Envelopes of Acceptability

2.0 Three Degree-of-Freedom Simulation Results

In summarizing the results obtained from a three degree-of-freedom simulation, the longitudinal system response characteristics are compared with the NADC criterion using:

(a) series or parallel servo modes,

(b) normalized pilot seat response, N_{z_p}, to step inputs of stick force command in the stability augmentation (SAS) mode, and

(c) normalized pilot seat response, N_{z_p}, for step inputs of normal acceleration command from the terrain following radar.

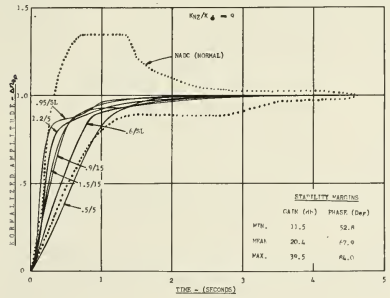
2.1 Stability Augmentation System (SAS) and Automatic Terrain Following System (ATFS).

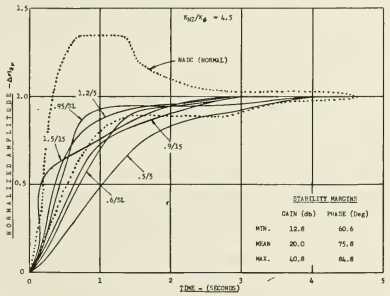
Early in the study, it was found that an increase in the AFCS speed of response was needed to attain sufficient augmented aircraft performance improvement at the higher altitude flight conditions to meet the NADC specification requirements. Changes to the SAS and ATFS compatible with AFCS hardware were made. These changes included a modified forward loop compensation network, a shorter canceller time constant, the addition of a normal accelerometer structural filter, and an increase in the normal acceleration (N_z) to pitch rate (θ) feedback gain ratio. The original AFCS design utilized an N_z to θ feedback gain ratio of 3:1 in units of g's/deg/sec. In the course of the McDonnell analytical studies, system configurations were studied using N_z to θ feedback ratios of 3:1, 4.5:1, and 9:1. A comparison of the responses for the ATF parallel mode with the NADC criteria are shown in Figure 6. As can be seen, the system with the 9:1 gain ratio provided the best system responses for meeting the criterion envelope. The stability margins are considered satisfactory for all three system configurations.

Results obtained from both the subsequent six-degree-of-freedom simulation study, incorporating 666A AFCS hardware and the fixed base flight simulator, and from flight testing indicated that the final system design should be changed to incorporate the 4.5:1 ratio. This change was necessary due to compatibility problems which existed between the force balance artificial longitudinal feel system used in the F-4 aircraft and the integrated servo-actuator in the parallel mode configuration.

Stability augmentation system responses to step

inputs of stick force command for both the parallel and series servo configurations are compared with the NADC criterion in Figure 7 for the final system design incorporating the 4.5:1 N_z to θ gain ratio. As can be seen, the normalized pilot seat normal acceleration responses for the high "q" LAHS flight conditions studied fall within the NADC envelope of acceptability. The responses shown for the lower "q" flight conditions are considered satisfactory with the use of a fixed-gain system. Of course, some compromise in response time at the low "q" and high altitude flight conditions had to be made using





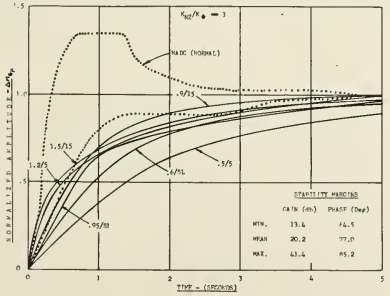
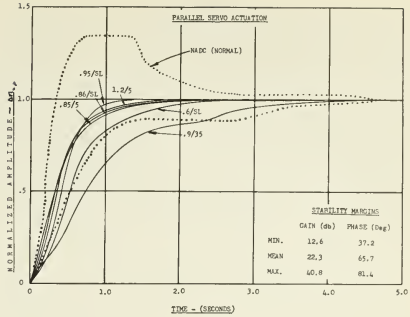


Figure 6 - Comparison of Normalized ATFS Pilot Seat Acceleration Time History Responses to Simulated Radar Step Command Input with NADC Criteria - Parallel Servo Mode
Weight = 37,200 lbs., c.g. = 30.8% $\bar{c}$



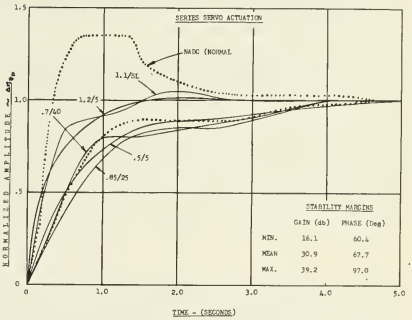


Figure 7 - Comparison of Normalized SAS Pilot Seat Acceleration Time History Responses to Unit Step Force Command Input with NADC Criteria
(K_{nz}/K_θ = 4.5)
Weight = 37,200 lbs.; c.g. = 30.8% $\bar{c}$

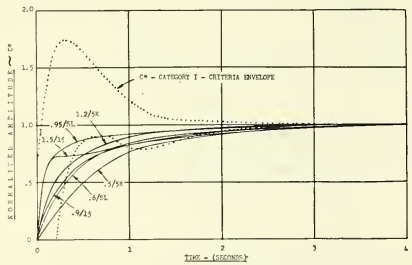
a fixed-gain system in lieu of a self-adaptive gain changing or air data gain-scheduled system. However, this compromise is considered justified when this slight reduction in performance is weighed against the increase in system reliability and decrease in complexity and maintenance time gained by deletion of the three-channel gain-changing function. Since for this program, the primary mission usage of the AFCS is in the low altitude, higher "q" regions compatible with automatic terrain following flight, the design of optimum system performance was tailored to this portion of the F-4 flight envelope. The stability margins for the SAS mode are considered adequate to insure stability in the presence of hardware tolerance variations and nonlinearities.

Responses for the final system design to simulated radar step command inputs in the automatic terrain following mode for the parallel servo configuration are compared with both the NADC criterion and C* criterion in Figure 8. Due to the concept similarity of the two criteria, similar evaluation results are obtained from comparison of the normalized time history responses with the two criteria as shown in Figure 8. The fixed gain ATF mode provides acceptable transient responses for all the LAHS flight conditions.

67-569

2.2 Attitude and Altitude Hold. The dynamic characteristics and stability of the outer loop modes of attitude and altitude hold were also studied in this phase of the program. Improvements to the pilot relief modes include a decreased attitude synch drive speed and the addition of an altitude rate signal for additional compensation. The normal accelerometer feedback signal is removed during outer loop operation and an "Up Stabilator" signal is added to improve attitude and altitude hold mode characteristics for turning flight. This signal is scheduled as a function of velocity and bank angle and is used to cancel the steady state pitch rate signal present in a turn. Only lateral stick motion is required to command a coordinated turn for control stick steering (CSS) with the outer loop modes of either attitude or altitude hold mode engaged. The performance of the attitude and altitude hold modes as studied was found to be much improved over that obtainable with the production F-4 AFCS, and both modes yielded adequate stability margins at all flight conditions studied.

2.3 Stick Force Per "g" Characteristics. The 666A AFCS was designed to meet the stick force per "g" specification requirements of MIL-F-8785. The stick force gradient, shown plotted versus aircraft velocity in Figure 9, meets the specification requirements. As a result of the use of integration in the inner loop for both the parallel and series servo configurations, the stick force gradient of



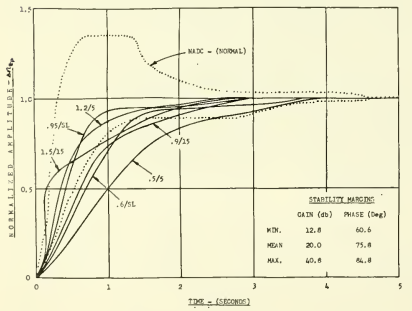


Figure 8 - Comparison of Normalized ATFS Pilot Seat Acceleration Time History Responses to Simulated Radar Step Command Input with NADC and C* Criteria, Parallel Servo Mode ($K_{nz}/K_\theta = 4.5$) Weight = 37,200 lbs.; c.g. = 30.8% $\bar{c}$

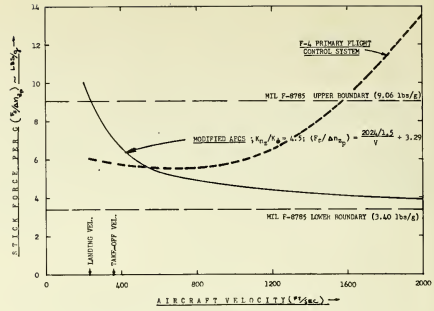


Figure 9 - Stick Force Per "g" Gradient Characteristics

the basic aircraft is very nearly masked. The extent to which this masking is possible during a maneuver is dependent on the deflection authority and follow-up trim motor rate allowed in the series servo mode and on the servo stall torque and the amount of mismatch between the stick force gradient of the basic aircraft and the stick gradient desired for the control augmentation system in the parallel mode configuration. In the F-4 aircraft, this mismatch notably increases with increasing Mach number due to the spring constant characteristics of the bellows in the force balance system which produces forces proportional to dynamic pressure.

2.4 Stability Margins. A notable result was obtained from this analysis as related to system stability margins. Referring to Figure 10, it can be seen that the gain margins for the final system design are low for the high "q" regions and high for the low to medium "q" regions, whereas the phase margins manifest an opposite characteristic.

The gain margin trend is readily explainable. In a self-adaptive or air-data gain-scheduled system, the forward loop gain is very nearly inversely proportional to the surface effectiveness parameter. With a fixed-gain system, the gain is limited to the maximum value that can be tolerated for stability reasons, i.e., where the surface effectiveness is the highest, occurring generally in the

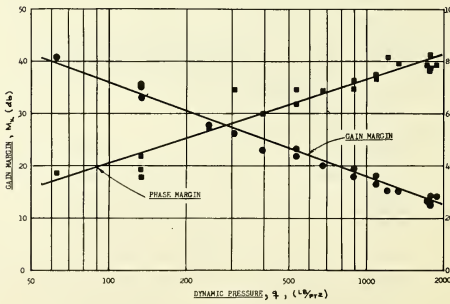


Figure 10 - SAS Stability Margins - Parallel Servo Mode - ($K_{nz}/K_\theta = 4.5$)

high "q" region. Consequently, the gain margins with a fixed-gain system are maximum for the low to medium "q" region where the surface effectiveness parameter is low.

The phase margins are observed to be lowest in the low to medium "q" region. The AFCS system dynamics in this portion of the flight envelope of the F-4 aircraft generate root loci branches that are nearer to the origin and closer to the imaginary axis than those same branches generated for the high "q" region. This characteristic is typical of stability-augmented aircraft in the high speed, fighter/interceptor category, such as the F-4. With a fixed-gain system, where large gain margins exist in the low "q" flight regime, the phase margins are even lower than with gain-scheduled systems. This characteristic is due to the fact that the phase margins are a function of the system gain and the proximity of this root locus branch to the imaginary axis. The phase margins approach values which are very nearly determined by the location of the basic aircraft poles in this portion of the flight regime. Therefore, the phase margins are minimum at the low "q" region and maximum at the high "q" regions as depicted in Figure 10.

3.0 Six Degree-of-Freedom Analog Simulation Results.

In addition to the three degree-of-freedom analysis, a six degree-of-freedom large perturbation man-in-the-loop, terrain following analog simulation was performed.

The purposes for this simulation were:

- o Pilot evaluation of all AFCS control and pilot assist modes using AFCS hardware,
- o Pilot evaluation and documentation of terrain following performance for the AFCS/Radar/Airframe combination under manual and automatic control, and
- o Limited evaluation of simulated system failures.

3.1 Description of Simulation. A functional block diagram of this simulation is presented in Figure 11. This simulation incorporated a six-degree-of-freedom airframe, AFCS hardware, a radar simulator, and a fixed-base flight simulator.

3.2 Terrain Following Performance. Typical terrain following performance achieved utilizing the simulated radar and hardware AFCS is shown in Figure 12. Both automatic and manual terrain following performances for the .9 Mach number flight condition are illustrated.

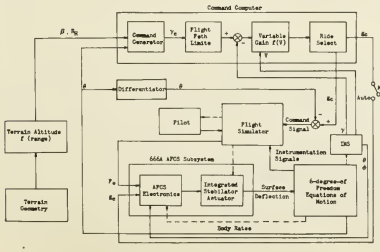


Figure 11-Terrain Following Simulation Functional Block Diagram

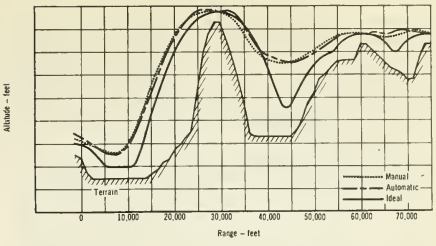


Figure 12 - Terrain Following Performance

It is noted that very little difference exists between the manual and automatic terrain following performance and both are considered excellent. McDonnell has determined that, for short duration flights, the human pilot utilizing an optimized command display can equal the performance of an automatic system. However, as the time duration of terrain following flight increases and/or other cockpit duties are imposed, the human pilot performance becomes grossly degraded.

The third profile shown in Figure 12 has been called the "ideal profile". This curve represents the best possible terrain following performance that can be achieved at .9 Mach number within the constraints imposed on the aircraft for the terrain following mode, i.e., limitations on load factor, climb and dive angles, and level flight over peaks and flat terrain.

3.3 Evaluation of Failure Effects. As a part of the six-degree-of-freedom simulation, a failure analysis was conducted to examine the effect of failures which were basic to the aircraft, the AFCS, and terrain following radar. These random failures were imposed upon the aircraft while pilots were "flying" the flight simulator at low altitude flight conditions. The response of the airframe to the failures and the ability of the pilots to take corrective action quickly enough to avoid collision with the ground was evaluated. The results of these tests showed that the pilots experienced little difficulty in detecting and correcting for failures which were indicated by the illumination of a failure warning light. However, certain failures which did not result in the illumination of a failure warning light went undetected, and collision with the simulated terrain occurred.

As an example, one of the non-indicating failures which can occur is an electrical open or hard-over on the normal acceleration command (g_c) output of the terrain following computer. No failure detection logic is provided for this non-redundant circuit. The only means of failure detection for the pilot is the monitoring of the "E²" situation display and the flight instruments. A typical analog trace representing the results of a hardover failure in the TF command signal and resultant ground impact is presented in Figure 13. While modification of the radar computer was outside the scope of this program, all such potential failures were documented for Air Force review.

IV. Flight Test Results

1.0 Description of Flight Test Program

Flight test evaluation of the 666A AFCS was

initiated on 6 October 1966 and completed in December 1966. Only seventeen flights were required to evaluate the AFCS. All flight testing was accomplished at St. Louis. Further flight test development and evaluation of the automatic terrain following system will be performed by the Air Force at WPAFB following the McDonnell program.

1.1 AFCS Flight Evaluation. The test pilots who flew the aircraft during the AFCS flight test program carefully evaluated the characteristics of the stability (control) augmentation system. This concept of augmented control would be utilized as the normal control system of the aircraft if this system were used for a production aircraft. Reversion to the present mechanical-hydraulic control system would be used for back-up if failure of the SAS occurs. Therefore, special attention was given to evaluating the handling qualities provided by the SAS concept of control.

The performance of the outer loop modes of operation including attitude and altitude hold was also evaluated during the flight test program. Simulated radar step command inputs were used to verify proper terrain following mode operation.

The results of these flight tests were compared with those of the six-degree-of-freedom simulation,

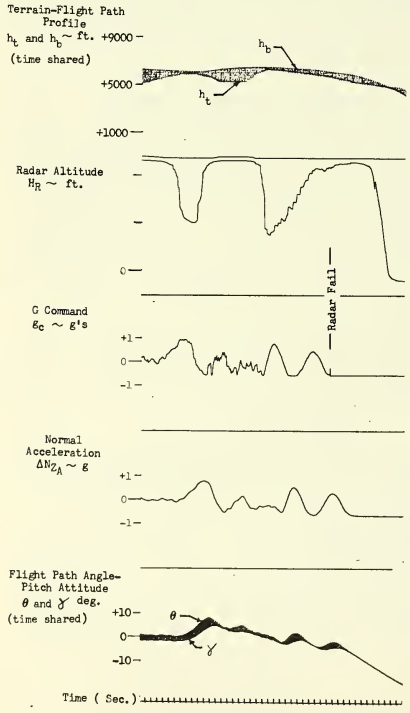


Figure 13 - Terrain Following Computer Simulated Hardover Failure - ATF Mode

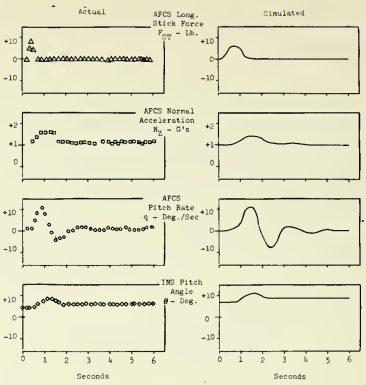


Figure 14 - Stability Augmentation System, Series Servo, Flight Test and Simulation Comparison Plot M = .5, h = 10,000 ft.

and the results of representative parameter comparisons are shown in Figures 14 through 17. The aircraft response to a sudden release of the control stick during a pull-up with the stability (control) augmentation mode engaged and the series actuator mode selected is presented in Figure 14. Figure 15 depicts the same mode and maneuver with the parallel actuator mode selected. The results of a maximum pull-up command with the terrain following mode engaged and the parallel actuator selected are presented in Figure 16. The aircraft response following a stick release during a pull-up with automatic engagement of pitch attitude hold with the series servo configuration is presented in Figure 17. In each case, very close correlation between the simulation and flight test results has been achieved.

From such comparisons with flight test recordings and from pilot comments, these significant

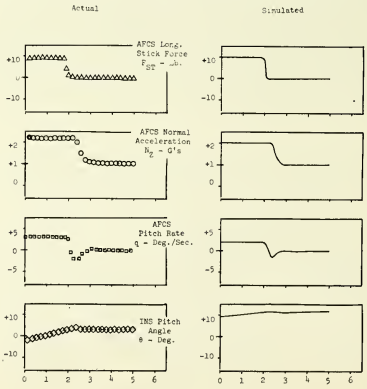


Figure 15 - Stability Augmentation System, Parallel Servo, Flight Test and Simulation Comparison Plot M = .95, h = 5000 ft.

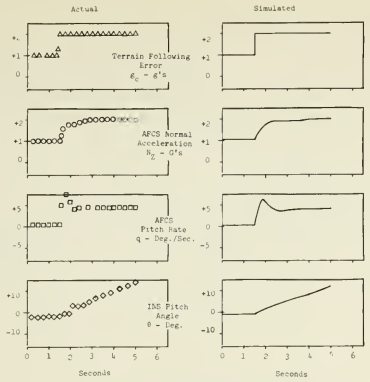


Figure 16 - Automatic Terrain Following, Parallel Servo, Flight Test and Simulation Comparison Plot M = .95, h = 5000 ft.

results were realized:

(1) Aircraft response during maneuvering in the SAS mode with the pitch actuator in the series or parallel configuration was considered satisfactory with deadbeat damping at all flight conditions.

(2) Overall performance obtained with the pitch actuator in the parallel configuration was considered superior in both the SAS and ATF modes. The series servo deflection limit and the low automatic trim rate used in the F-4 aircraft resulted in a lower pilot opinion rating of the series servo operation.

(3) Aircraft response to the 2 "g" normal acceleration command with the system in the ATF mode was demonstrated satisfactorily during operation in both pitch actuator servo modes.

(4) Altitude hold mode performance was successfully demonstrated throughout the aircraft flight regime.

(5) Pilots confirmed their preference for nearly constant stick force per "g" handling characteristics throughout the flight envelope.

(6) The near-zero stick-force-per-V handling characteristics were found to be acceptable by the pilots.

(7) Pilot comments indicated that it was difficult to discern improvement in aircraft handling qualities provided by the stability (control) augmentation system at the low "q" flight conditions due to the large stick deflections which were required for commanding maneuvers.

In addition, a number of modifications to the system were accomplished in the process of performing these flight tests, and several areas of future improvements were defined. Major items are summarized as follows:

(1) As originally configured, the pitch stability augmentation mode could not be used for takeoff due to the operation of the automatic trim follow-up in the series mode and due to the main-ram follow-up canceller driving the surface hard-over in the parallel mode when stick forces were applied. Since the series damping mode is desired for takeoff, modifications were made to disable series auto-trim when weight is on the main gear in order to provide for automatic removal of the trimming function when the aircraft is on the ground. Time limitations precluded incorporation of feasible changes to the hardware to facilitate takeoff with the system using the parallel servo configuration.

(2) With the stability (control) augmentation system engaged and the series actuator selected, application of stick force by the pilot resulted in adverse stick force and stick motion due to the series automatic trim. This condition was eliminated by making the trim inoperative when stick force is applied. This change resulted in the series actuator bottoming whenever large stabilator deflections are commanded. No increase in series servo authority was made to alleviate this problem during this program.

(3) Pitch and roll attitude hold operation was satisfactory for straight and level flight and during maneuvering at small bank angles. Normal acceleration transients occurring during pilot commanded pitch maneuvers performed in turning flight were eliminated by varying the outer-loop fader time constant as a function of bank angle.

(4) The use of limited authority series servo and automatic trim follow-up employing slow trim rates significantly increases the system response times for large input commands. Faster response times can be obtained using either a parallel servo or a series servo with large authority.

(5) Significant reduction in stick motoring was achieved by lowering the feedback gain ratio. The small level of stick motoring or "nervousness" that still exists in the parallel servo mode was considered objectionable for precise manual terrain following.

(6) Small pitch attitude changes were difficult to accomplish in the series servo mode due to a combination of low frictional forces in the control system and a relatively slow force buildup from the automatic trim. This condition could be improved by lowering the force breakout level in the stick force transducer. Since this adjustment for the triply redundant transducer was a factory adjustment, this modification was not evaluated during the flight test program.

(7) The lack of a manual pitch trim capability for trimming the aft stick force during a turn with

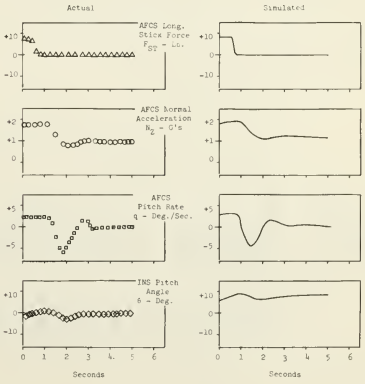


Figure 17 - Attitude Hold, Series Servo, Flight Test and Simulation Comparison Plot M = .5, h = 10,000 ft.

the stability (control) augmentation system engaged was considered unacceptable. Consideration was given to installing an integrating amplifier, which could be controlled by a manually actuated trim switch, to provide this capability. The output of this amplifier would be summed into the AFCS bridge to provide a voltage equivalent to the output of the stick force transducer. In this manner, the pilot force can be relieved to zero during a turn. However, this modification could not be readily made in the AFCS hardware and was therefore not accomplished.

(8) Poor stall recovery characteristics were present in both pitch actuator modes. In series mode, the stick remains fixed following a stall condition. In parallel mode, the servo moves the stick aft to oppose the nose down pitch rate following a stall. This problem is inherent in all stability (control) augmentation systems of this type. No attempt was made during the flight test program to incorporate system changes to alleviate this condition.

Since these undesirable characteristics of the AFCS do remain, further development would be necessary before implementing it as a production system. On the other hand, the system as it exists at the completion of the flight test program is considered suitable for development test flying of automatic terrain following systems. This system will provide satisfactory flying qualities and the necessary reliability for low altitude high speed flight.

V. Conclusions and Recommendations

Engineering knowledge and technology have been enhanced as a result of the 666A Program in the areas of advanced automatic flight control system synthesis techniques, handling qualities, and flight testing. The following major program conclusions were reached:

(1) Simulation and flight test results indicate that the 666A AFCS as it exists at the completion of the flight test program will provide the high performance and high degree of reliability required to accomplish the automatic terrain following task in the YRF-4C test aircraft.

(2) The 666A AFCS fixed-gain longitudinal channel performance and handling qualities satisfactorily meet the NADC and C* criteria for the low-altitude high-speed environment.

(3) The simulation studies:

- (a) significantly reduced the flight test development time,
- (b) were instrumental in determining and defining necessary AFCS modifications to enable the AFCS to meet the specific requirements, and
- (c) permitted pilot evaluation of all AFCS modes of operation, cockpit instruments and displays, and terrain following performance.

(4) The redundant solid state circuitry with electronic majority-voting logic provides undegraded system performance following the occurrence of an in-flight sensor or electronic system failure.

(5) The 666A integrated actuator hardware, which features failure detection and correction logic, provides high reliability for operation throughout the flight envelope.

(6) The characteristics of the existing aircraft primary flight control system and power supplies should be included in the early design phase of the AFCS to minimize the mechanical and electrical incompatibility and mismatch which could produce significant system operational problems.

The following recommendations are made:

(1) Failure analysis studies should be performed on automatic flight control systems operating in the LAHS environment. Adequate precautions must be taken in the system design stages to prevent situations where any one single malfunction in the aircraft power systems, the AFCS, or terrain following radar requires immediate and/or precise action by the crew to avoid disaster.

(2) Although the 666A limited authority series servo operation is considered sufficiently reliable for LAHS operation, it is recommended that future systems, which employ full authority series servo operation in the pitch axis, be made even more reliable. Hardover failures occurring with large authority series servos cannot be overpowered. Reliability must be very high to insure the necessary margin of safety to avoid disaster.

(3) In future aircraft designs incorporating combinations of mechanical backup control systems and electronic control augmentation systems, it is recommended that the mechanical system design possess the nearly constant stick force per "g" characteristics desired by the pilots. Such a design will facilitate the interfacing of the stability (control) augmentation system with the basic manual control system of the aircraft.

No. 67-570

THE SST FLIGHT CONTROL SYSTEM CONCEPT

by

JAMES R. HORSNELL
The Boeing Company
Seattle, Washington

AIAA Paper
No. 570

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

THE BOEING SST FLIGHT CONTROL SYSTEM CONCEPT

James R. Horsnell, Senior Group Engineer,
SST Flight Controls Group,
The Boeing Company, SST Branch,
Seattle, Washington

Abstract

The Boeing 2707 commercial Supersonic Transport airplane, while performing throughout a much larger flight envelope with no increase in required flight crew skill levels, requires a Flight Control System with reliability equal to or exceeding that of existing subsonic transports. Reliability for safe operation is assigned top priority, while reliability to assure scheduled operation is recognized as an economic necessity. Redundancy is employed in the form of multiple hydraulic power systems, surface actuators, mechanical load paths, and electronic subsystems. A comparison of control surface hinge moments with those of subsonic jets dictates a fully powered actuation system with no direct pilot manual reversion. Long mechanical control system runs require the incorporation of pilot assist servos to obtain surface resolution. In order to reduce flight crew work load and attain the increased system performance requirements, utilization of electronic control subsystems is increased. An electric command subsystem is combined with a mechanical system for optimum performance and reliability. Hydraulic and electronic systems are designed to remain operational following the failure of one system and to provide safe flight following failure of two systems. System design is verified in exhaustive test programs.

I. Introduction

The basic task of the 2707 Boeing SST (Figure 1) is to increase productivity, in terms of payload/miles, over that of subsonic transports while retaining or improving upon their economy and safety levels. Principal areas of difference between the Boeing SST and current subsonic transports which influence flight control system design are:

(a) a larger speed range up to M 2.7
(b) larger gross weight and mass moments of inertia (pitch moment of inertia is approximately ten times that of the 707-320)
(c) a large range of environmental temperatures with maximums over 400°F due to aerodynamic heating
(d) reduced thickness ratio fixed and moveable flight surfaces dictated by aerodynamic drag
(e) variable sweep wing geometry for advantages in aerodynamic performance

The flight control system must provide airplane control and handling qualities as good as those of current subsonic jets with no increase in flight crew skill levels. Although an increase in automatic and electronic flight control modes is indicated to minimize flight crew work load and supplement their inherent capabilities, the pilot shall have the capability of overriding all such modes. This shall be accomplished as much as possible by utilizing his natural reactions.

Trailing-edge flight control surfaces are utilized on the fixed horizontal stabilizer and on the moveable wings for roll and pitch control. Lift reducing spoilers are also used to supplement roll control. A multiple segmented rudder is employed for directional control. With moveable wings swept aft for supersonic flight, the resulting mating with the horizontal stabilizer produces one continuous airfoil section. When the wings are swept forward for subsonic cruise and low speed landing and takeoff conditions, trailing-edge flaps and leading-edge slats are extended.

II. Reliability

Reliability to assure safe operation is assigned top priority in Boeing SST design, while reliability to assure continuous on-schedule operation is recognized as an economic necessity.

Figure 2 compares dispatch reliability histories of 707 and 727 airplanes to 727 and SST goals. Based upon reliability analysis to date, the 99% goal for Boeing SST dispatch reliability appears achievable.

During the 727 design stage, a comparison of relative subsystem complexities of the 727 and 707 was made and flight delay factors established. These factors were applied to 707 historical delay experience to predict 727 dispatch reliability. In addition, standard mathematical probability formulas were used to predict 727 component failure frequencies on the basis of failure histories of similar equipment on earlier Boeing jet transports and military aircraft. These frequencies were then used to predict 727 dispatch reliability. Where indicated, redesign followed.

FIGURE 1 THE BOEING 2707 SUPERSONIC TRANSPORT

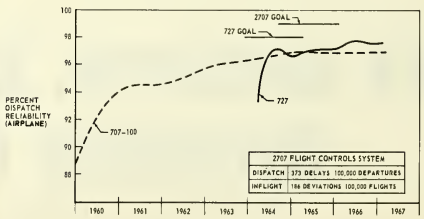


FIGURE 2 RELIABILITY

Boeing SST reliability efforts employ these same techniques plus a failure mode and effect analysis to determine the effects of all failures and to generate redesign efforts. The goal is to eliminate those factors which adversely affect safety and dispatch or in-flight reliability. Computerized reliability calculation methods are employed in trade studies of various subsystem design concepts. The increased emphasis on maintainability is also supported by failure mode and effect analyses.

The flight control system reliability allocations noted in Figure 2 were established by first distributing airplane reliability goals to all systems in proportion to achieved reliabilities of corresponding subsystems on earlier Boeing transports. Suitable adjustments were made to reflect differences of configuration, usage, and environment of the SST. Following system reliability analyses, readjustments in allocations were made to meet overall airplane goals.

The degrees of redundancy established for the flight control system are dictated by safety criteria, dispatch reliability requirements, and in-flight reliability requirements, in that order of priority. In general, all channels in redundant systems must be operative for dispatch; however, some automatic flight control modes are not considered essential to safe flight in all cases. For example, automatic all-weather landing mode redundancies could be reduced given favorable weather conditions.

The 727 history indicates that operational factors exerted the major influence upon attainment of dispatch reliability goals. 727 fleet reliability referenced in Figure 2 is influenced greatly by individual operator's maintenance practices, route schedules, reserve aircraft status, parts inventories, etc. In this regard, dispatch reliability is affected critically by the ability to perform unscheduled maintenance in the allocated stop-over or turn-around times. In order to minimize such unscheduled maintenance and speed up the action when necessary, considerable emphasis is being placed upon built-in test equipment and airborne integrated data systems for the Boeing SST.

III. System Design Features

Hydraulic System

Figure 3 illustrates the 2707 hydraulic system. A significant feature is the simultaneous operation of three completely independent 3000 PSI primary systems to eliminate the need for shuttle valves or pilot switching should one system fail. All three primary systems supply power to the primary flight controls and two of these systems are used exclusively for this purpose.

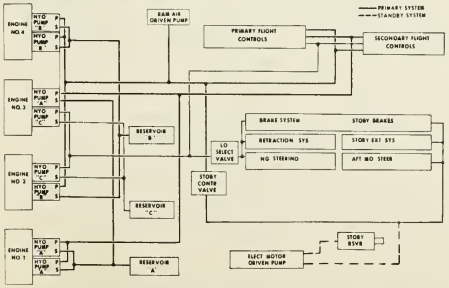


FIGURE 3 HYDRAULIC SYSTEM DIAGRAM

Engine-driven pumps are located such that failure of any one engine or failure of two engines on one side will not result in loss of any hydraulic system, and failure of any two engines will not result in loss of more than one hydraulic system. The air-driven turbine supplies adequate hydraulic power for safe flight control and landing in the event of failure of four engines.

Figure 4 emphasizes the fact that the flight control system sizes the 2707's hydraulic systems, whereas previous subsonic jet hydraulic systems have usually been sized by landing gear requirements. SST hydraulic flow requirements are dramatically larger than those of the 707-320. Total flow capacity of the 2707 is in excess of 1000 GPM — compared to 46 GPM on the 707-320.

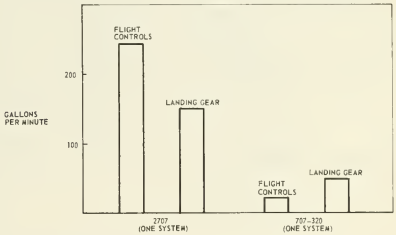


FIGURE 4 MAXIMUM HYDRAULIC LOAD

Master Servos

As shown in Figure 5 a master servo or pilot assist servo is placed in the mechanical cable system between the pilot's control and surface actuators. Long cable runs with their attendant friction and hysteresis do not lend themselves to attaining precise force and positioning accuracy.

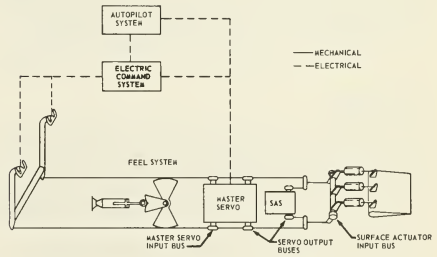


FIGURE 5 PRIMARY FLIGHT CONTROL SCHEMATIC

The master servo receives electric command signals generated by control forces and amplifies these signals hydro-mechanically to drive cables and linkages controlling the surface actuator valves. Figure 6 illustrates one of the master servo concepts being tested and evaluated. Three separate housings are bolted together to form a triple channel power unit with each channel powered by a separate hydraulic system. One of these channels is shown in the section view. A connector assembly facilitates removal and replacement of the power unit. Electric signals from pilot force transducers or from the autopilot system normally control the three torque motors which actuate the three parallel jet-type valves. The

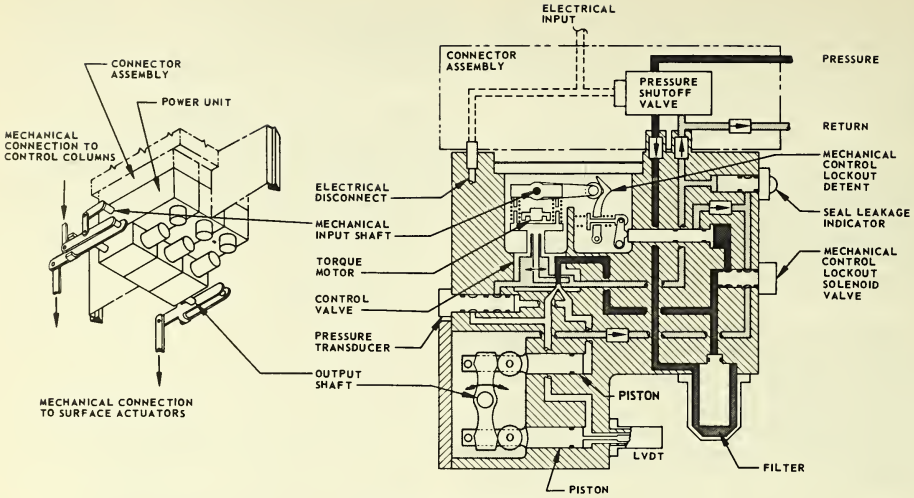


FIGURE 6 MASTER SERVO

three sets of actuator pistons are force summed on the output shaft. The mechanical connection to the control column repositions the column or wheel. This mechanical connection may also be utilized to introduce mechanical control inputs to the three valves by overpowering the mechanical control lockout detent. A pressure transducer in each channel is used to compensate electrical input signals and thereby equalize actuator loads. In event of loss or deactivation of the Electric Command System, mechanical inputs are introduced directly to the three control valves and the mechanical control detent is deactivated.

Dual Load Path

Figure 5 illustrates the basic 2707 dual load path design approach. It is a design objective that all mechanical load paths will incorporate at least one level of redundancy. Each of the dual load paths shall be designed for full strength capability. In order to realize the maximum benefit from this redundancy, interconnecting bus shafts will incorporate force limiters as illustrated in Figure 7. In this example, an obstruction in either cable system will permit servo outputs through the other system. If the forces on the discrepant cable system return to acceptable limits, the force limiter re-engages.

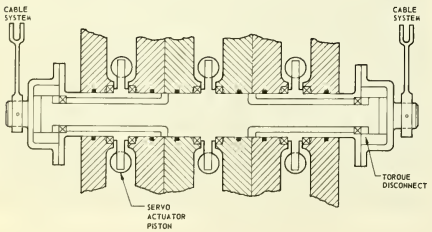


FIGURE 7 SERVO OUTPUT BUS

Surface Actuators

The 2707 primary surface actuators are required to generate significantly higher hinge moments than subsonic jets. (Elevator hinge moments are approximately twenty-five times those of 707-320 elevators).

The fact that these surfaces cannot be balanced for flutter or transonic buzz dictates that each of the primary flight control surfaces be powered by three independent hydraulic systems utilizing parallel actuators force summed on each surface. A typical actuator design is illustrated schematically by Figure 8. Hydraulic and electrical connections are introduced through a connector assembly containing an electrically operated shut-off valve and a check valve in the return line. When the actuator is removed from its attach points the pressure shut-off valve is closed. Fluid leakage from the actuator is prevented by a check valve in the pressure line and a low pressure relief valve in the return line. The main control valve incorporates an anti-jam detent and dumping valve such that if the valve jams, the input lever motion compresses a preloaded spring, opening a poppet valve which releases main system pressure. This pressure drop actuates the pressure operated by-pass valve, thus connecting both sides of the main actuator piston. In this manner the remaining parallel actuators may move the control surface without resistance from the discrepant actuator. The actuator piston rod utilizes one high pressure and one low pressure seal, with the cavity between ported to return via a seal leakage indicator. Excessive leakage causes an increase in pressure in the cavity which actuates the visual indicator.

Trim System

Figure 9 illustrates the significant feature of the primary flight control trim system. Differential output

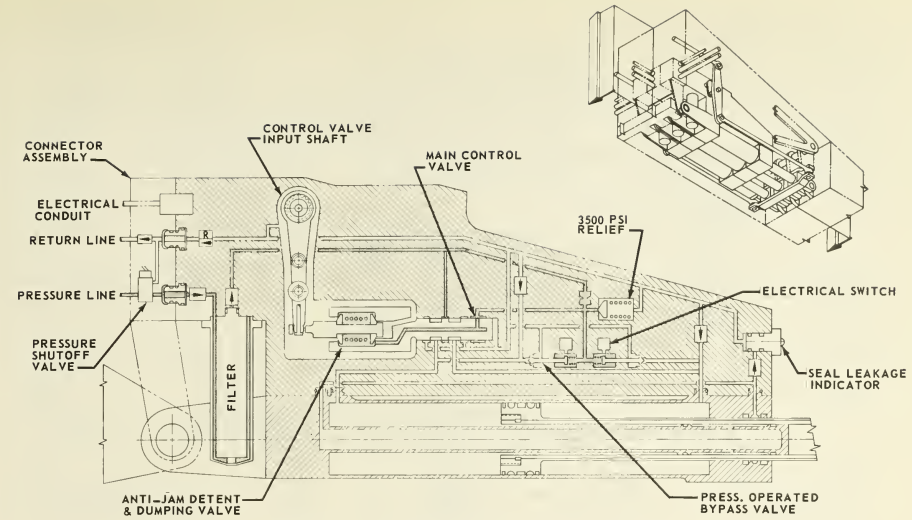


FIGURE 8 FLIGHT CONTROL SURFACE ACTUATOR

of the dual transducers is converted into a signal for trim rate and direction which controls a nose-up or nose-down pulse generator and a ground-path gate generator. The stepping motor responds only to properly sequenced pulses and malfunctions do not cause unscheduled motor operation.

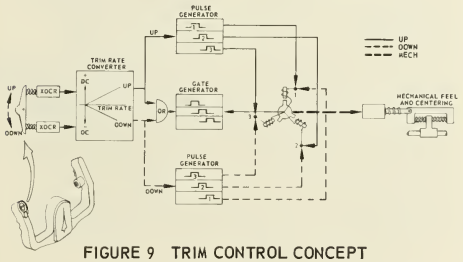


FIGURE 9 TRIM CONTROL CONCEPT

Figure 10 shows this concept incorporated into the control system. Trim is introduced in a similar manner in all three axes, and the electric command system and mechanical backup system trim is synchronized. Electric trim circuits and motor windings are dualized to provide separate circuits for pilot and copilot. Autopilot and SAS off-load trim inputs are introduced using these two circuits plus a third circuit and a third motor winding.

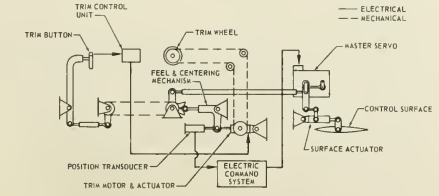


FIGURE 10 TRIM SYSTEM

Flight Control Electronics

Increased performance requirements of the 2707 entail greater dependence upon electronic subsystems to optimize control and handling qualities without increasing flight crew work load. Multiple redundancy is employed to insure flight safety and reliability. Proven state-of-the-art design concepts are utilized. Extensive flight simulator and laboratory testing is underway for developing and qualifying the subsystems.

The Automatic Flight Control System (Figure 11) includes three channels for autopilot; three channels of stability augmentation system for the roll axis and four for pitch and yaw axes; and two channels of autothrottle. These provide pilot relief functions, all-weather landing capability, improved airplane stability and handling qualities, and automatic thrust control.

Most of the autopilot modes are conventional; however, the climb and descent modes provide capability of holding the airplane to a computed Mach-altitude schedule to provide a selected sonic overpressure.

The present Stability Augmentation System does not directly attempt to dampen body bending modes. However, mode suppression is being examined to improve both ride comfort and structural life.

The Electric Command System supplements the Mechanical Control System by providing optimum manual control forces and system resolution (Figure 5). Pilot manual control forces are converted to electrical signals which amplify and gain schedule these commands and assure precise control surface positioning. Variable gain scheduling controlled by airspeed data provides artificial feel in the electric command mode. The

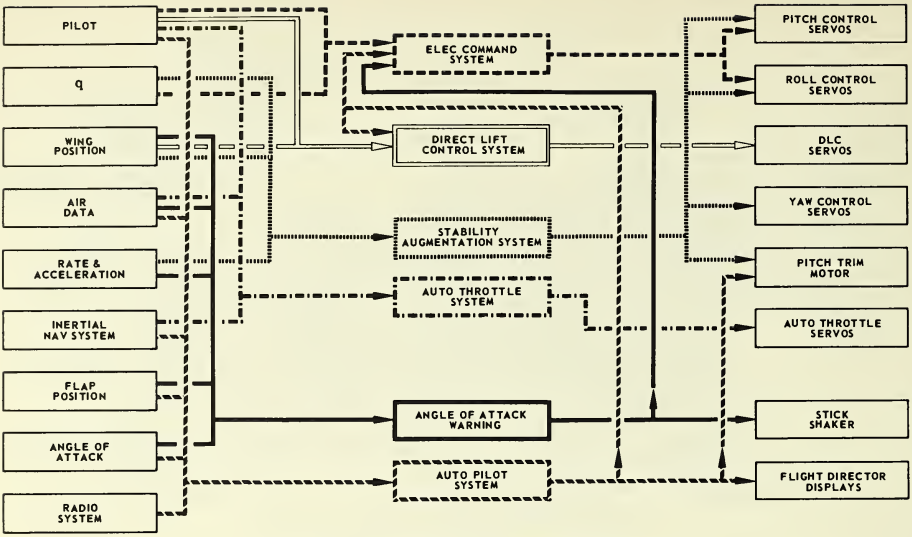


FIGURE 11 FLIGHT CONTROL ELECTRONICS

mechanical feel and centering mechanism shown is utilized during the backup mechanical mode only. The Electric Command System incorporates three redundant electronic channels.

The Angle-of-Attack Warning and Control System prevents extreme pitch attitudes by sensing pitch acceleration and angle-of-attack and actuating a stick shaker to warn the pilot of excessive pitch-up rates or angle-of-attack being approached. If the pilot's response is not adequate a pitch-down signal is introduced into the Electric Command System. Two independent electronic channels are incorporated in the sensing and warning circuits and three in the pitch command circuits.

The Direct Lift Control System commands symmetrical modulation of wing spoilers in parallel with pitch commands to provide additional lift without increasing pitch attitude. Three redundant electronic channels are utilized.

Malfunction Detection

Figure 12 shows the electric command system to illustrate the malfunction detection and control concept employed. Redundant force transducer signals command parallel computers. The intermediate value of the three signals is selected by the mid-value logic circuit. Each input signal is continuously compared to this mid value and equalized to prevent tolerance build-up. An input error larger than the equalization limit will ground that channel and reduce its output to zero. Each computer output now equals the middle value of the zero signal and the two active signals and the system remains operational. In the event of failure of a second channel, the mid-value logic output becomes zero for all three computers, attaining the fail-passive objective. In a similar manner, the servo loops are monitored and equalized, the output load pressure of each servo being compared to the middle value of the three servos. The resulting error signals are used to equalize the three servos and to switch-out a discrepant servo when the equalization limit is exceeded.

Monitor and Control

The design philosophy for flight crew operational status displays and control is illustrated by Figure 12. Individual electronic channel or servo loop failures will illuminate the respective status lights located adjacent to the axis switch. The E/CS warning light and the system warning light on the master annunciator panel are also illuminated. Two electronic channel failures or two servo loop failures automatically disconnect the axis and illuminate the OFF light and the master warning annunciator lights. The pilot now has the option of engaging single channels. He may also elect to cycle the engage switch to OFF/RESET and back to AUTO. If the prior disconnect was due to a transient failure, automatic operation would resume. In the event of an axis disconnect, the master servo valves are controlled by mechanical cable inputs.

Additional data to isolate faults to the replaceable component level and to measure system and component degradation is also provided for maintenance personnel but not normally displayed to the flight crew.

Wing Sweep/High Lift System

The movement of wing trailing-edge flaps and leading-edge slats is coordinated with wing sweep angle to provide the most efficient geometry for the supersonic cruise, subsonic cruise, and low speed conditions. When the wing is aft, in the supersonic cruise position, flaps and slats are stowed to permit the wing to integrate with the horizontal tail and form one continuous airfoil section. As the wing sweeps forward to the subsonic cruise position, the trailing-edge flaps adjust to form an efficient wing airfoil section. Sweeping the wing farther forward and extending the flaps and slats permits landing attitudes comparable to those of fixed-wing subsonic airplanes. Coordinated movement of wing and high lift systems is

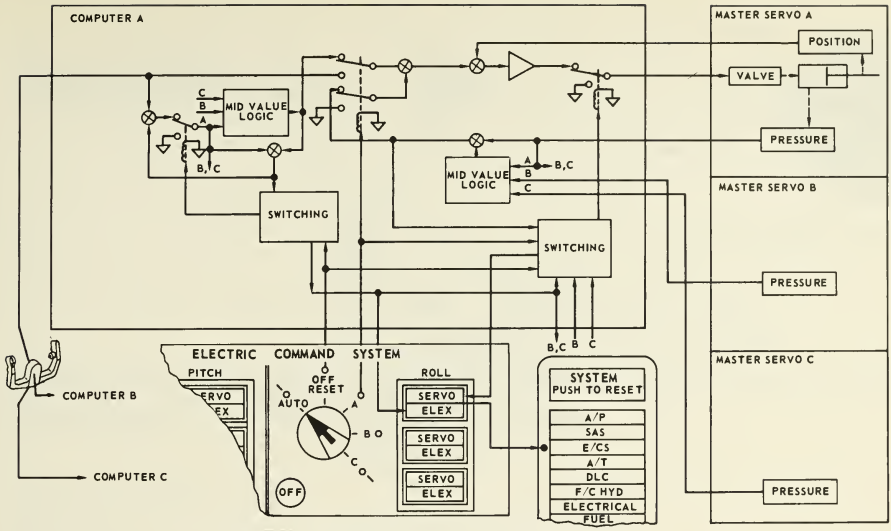


FIGURE 12 ELECTRIC COMMAND SYSTEM (ROLL AXIS)

controlled by a mechanical programmer. See Figure 13. Three primary hydraulic systems are utilized to provide actuation redundancy.

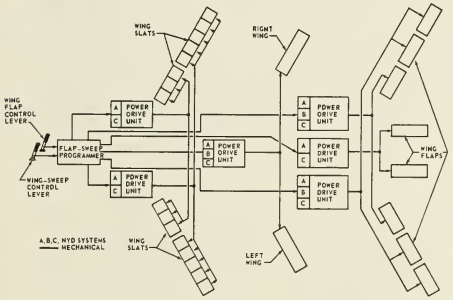


FIGURE 13 WING SWEEP/HIGH LIFT SYSTEMS

Wing sweep is achieved by the use of a pair of ball screw actuators. See Figures 14 and 15. Screws are synchronously driven by a central power unit through dual torque shafts. The dual shafts are geared together at each actuator so that a break or disconnect in one of the shafts will not interrupt power applied to both screws. The screws themselves incorporate dual load paths including a torsion-tension tube extending the length of the hollow screw to provide a dual tension load path and torsion drive path to allow continued operation even if a fracture occurs in the screw. The power unit is driven by three hydraulic motors, each capable of moving the wing through its full travel. Each motor, powered by a separate hydraulic system, drives one set of gears in the triple differential gear box. Loss of power to any motor, locks that motor shaft and permits power transmission from the active motor or motors.

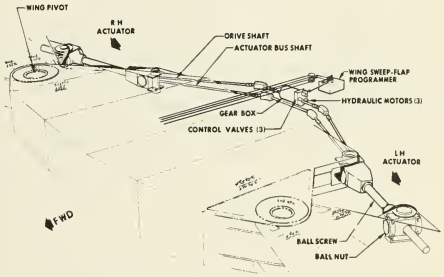


FIGURE 14 WING SWEEP SYSTEM

Direct Lift Control (DLC)

Precise flight path response characteristics during the landing approach are attained through Direct Lift Control which varies wing lift independently of angle of attack. This is done by modulating the wing spoilers about an initial spoilers-up trimmed configuration. Movement of wings forward to the landing position automatically raises the wing spoilers a few degrees and subsequent pitch control commands introduce signals to move the spoilers down to increase lift and up to decrease lift. This movement is in parallel with the normal pitch control surface movement. A steady state pitch command returns the spoilers to their initial DLC position. See Figure 16. The high mass moment of inertia of the 2707 tends to delay the airplane response to abrupt control inputs such as would be used for normal glide path control. Figure 17 shows the response to control inputs. The DLC provides SST flight path response similar to that of the smaller 707-320.

Temperature Effects

Higher environmental temperatures have an impact

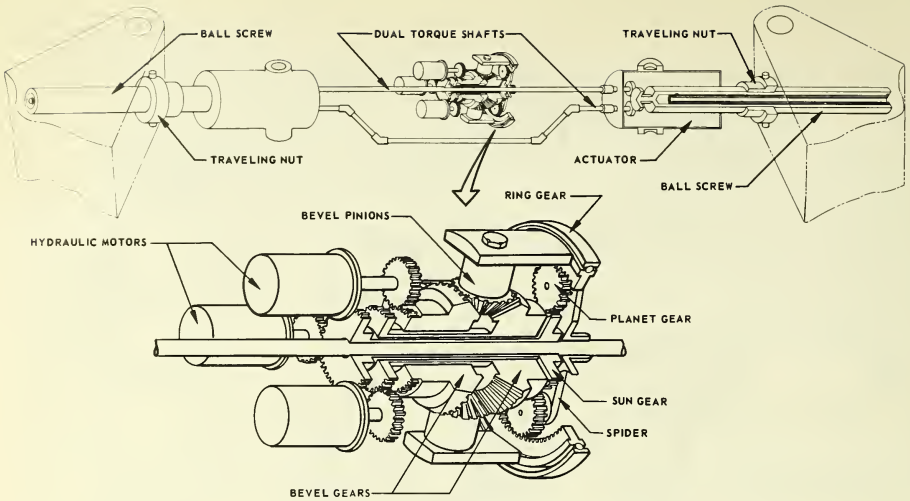


FIGURE 15 WING SWEEP ACTUATION

upon the SST flight control system design. The hydraulic system will utilize one of several high temperature fluids being tested. New hydraulic seals are being developed for swivels and actuators. Higher valve leakage rates will result from increased clearances due to thermal shock considerations. Special attention is being given to plain and anti-friction bearing design and solid-state lubricants are being evaluated. Consideration is being given to elimination of lubricants from lightly loaded anti-friction bearings. Steel and titanium alloys will replace aluminum in most applications over 250°F. No magnesium alloy usage is planned.

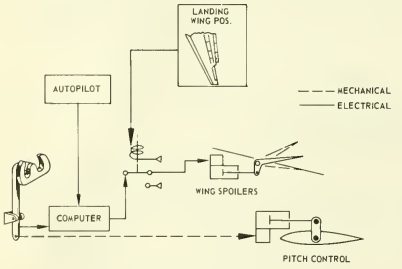


FIGURE 16 DIRECT LIFT CONTROL

IV. Test Programs

Major emphasis has been placed upon testing to develop and verify the Boeing SST design. In addition to those tests whose purpose is to develop design criteria and concepts, all critical components and subsystems are thoroughly tested to verify compliance with design objectives. Flight simulators are employed extensively to verify flight control system performance prior to first flight. A full-scale flight control servo simulator, similar to that used on Boeing 727 and 737 programs and equipped with actual airplane hardware, will verify flight control system response and stability. Developmental power servos have been procured and are currently being tested in manual and automatic control loops. As actual airplane control system hardware is developed, it will be placed in individual testing fixtures and evaluated statically and dynamically with simulated inertias, airloads and structural compliance under various environments. After each of the control surface subsystems has been successfully demonstrated, the several subsystems

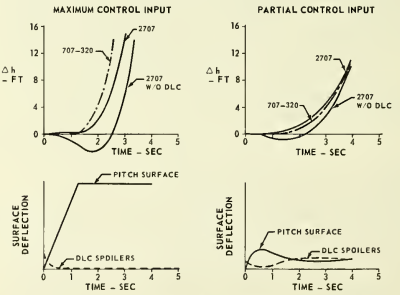


FIGURE 17 DIRECT LIFT CONTROL RESPONSE

will be integrated mechanically, hydraulically, and electrically and operated from a flight control cab to evaluate the complete control system response under all operating conditions. Analog computers will be used to simulate portions of the automatic flight control equipment, to simulate the airplane and its dynamics, and to generate control signals for the flight control system testing and evaluation. These tests will not only verify flight safety prior to first flight, but will reduce the scope of the flight test program.

No. 67-571

DESIGN OF ELASTIC MODE SUPPRESSION SYSTEMS FOR RIDE QUALITY IMPROVEMENT AND APPLICATION TO AN SST

by

LESTER D. EDINGER
Honeywell, Inc.
Minneapolis, Minnesota

AIAA Paper
No. 67-571

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.11, 7.08, 8.08

DESIGN OF ELASTIC MODE SUPPRESSION SYSTEMS FOR RIDE QUALITY IMPROVEMENT AND APPLICATION TO AN SST

Lester D. Edinger, Principal Engineer

Honeywell Inc., Aerospace Division
Aircraft Flight Systems Group
2600 Ridgway Road
Minneapolis, Minnesota

ABSTRACT

An analytical technique is presented for design of systems to actively control the structural bending modes to improve the ride quality on flexible aircraft. The technique specified a set of requirements for the sensor configuration and filtering for a system which provides corrective control through the conventional aerodynamic control surfaces. Vehicle data is required only in a frequency response format thus facilitating the inclusion of unsteady aerodynamics. Application of the technique to a large supersonic transport is described. Gust response data on the SST indicated that the lower frequency symmetric structural modes contributed up to 60 percent of the rms acceleration at the pilot's station. Results of the study showed that a system designed by the synthesis technique essentially achieved the performance objective of a 50 percent reduction in the rms acceleration level at the pilot's and other critical vehicle stations. This application of the synthesis technique to the SST problem showed it to be an effective design tool.

INTRODUCTION

In large aircraft such as the supersonic transport (SST) the ride quality deteriorates in turbulence due to the increased vehicle flexibility. To alleviate this problem the response amplitudes of the significant structural modes can be reduced through use of an active control system. The control system is designed to sense the structural motions and to generate a corrective control force through the aerodynamic control surfaces. To date, design of such a system (referred to as an elastic mode suppression system) has been accomplished primarily by trial and error techniques. The presence of factors such as unsteady aerodynamics and the relatively large number of parameters to be evaluated make these techniques cumbersome and limited in application. A synthesis technique has been developed which results in an orderly and systematic procedure for defining an elastic mode suppression system.

This paper describes the synthesis technique by first discussing the general mode suppression problem in terms of the design objectives and design approach considerations which form a basis for the development of an elastic mode suppression system. Using this information the formulas and conditions used in the synthesis procedure are established. Finally, use of the synthesis technique is demonstrated through an example using the SST with four symmetric structural modes.

The synthesis technique described herein was evolved from development work on a large supersonic transport. Some design considerations which aid in understanding the synthesis technique are drawn from the SST example. However, the technique is intended for a more general application and should not be construed as being restricted to this particular SST vehicle.

Application of the synthesis technique to the SST illustrated its effectiveness not only in deriving a solution to the problem but in other aspects. The performance results demonstrated that a significant degree of ride improvement can be attained through the use of an active control system. For the SST application a design goal of a 50 percent reduction in the rms acceleration levels in turbulence was essentially achieved at the critical vehicle stations.

THE GENERAL PROBLEM

Ride quality is measured in terms of the rms acceleration level occurring at various crew and/or passenger stations in the vehicle. Stations at which the acceleration levels are large in terms of human comfort ratings are termed "sensitive" stations. The acceleration experienced at these stations in air turbulence is a result of energy being transferred from the surrounding atmosphere to the vehicle not only through the rigid body motion but also through excitation of the vehicle's structural modes. The ride quality can be significantly improved over that of the unaugmented vehicle through the use of a rigid body gust alleviation system and/or an elastic mode suppression system. Only the design of an elastic mode suppression system is considered in this discussion. It is not the objective herein to develop a specific elastic mode suppression system (EMSS) but rather to describe a synthesis technique which can be used to define an elastic mode suppression system. This discussion will establish the design objectives and design approach upon which the EMSS design is based. Hence, these assumptions serve also as a basis for development of the synthesis technique.

Gust response data on vehicles such as the SST and XB-70 indicate that the lower frequency symmetric structural modes contribute significantly to the acceleration levels at the sensitive fuselage stations (e.g., the pilot's station and the aft passenger cabin). For example, in one of the configurations of the SST it was estimated that about 60 percent of the acceleration level experienced at a sensitive station could be

attributed to the first three structural bending modes, about 30 percent to the rigid body motion, and the remaining 10 percent to the higher frequency structural modes. Thus, it was evident that if the ride quality of the flexible vehicle was to be enhanced, an additional design objective was needed for the automatic flight control system. This design objective was to reduce the rms acceleration level at the sensitive fuselage stations by decreasing the response amplitudes of the significant structural modes (i. e. , usually the lower frequency modes). Uncertainties in the vehicle dynamics and increased system complexity make it unattractive to provide active mode suppression for the less significant higher frequency modes in order to alleviate the additional contribution which they make to the acceleration levels. Thus, it was an objective to minimize any coupling through the elastic mode suppression system with these higher modes.

Further, it was essential that satisfactory "rigid body" dynamics be maintained while decreasing the response amplitudes of the significant structural modes. This is important from the standpoint of assuring good pilot handling qualities and to prevent amplified turbulence response at the short-period frequencies.

An additional design objective was to establish an analytical design which allowed an economical and reliable implementation. System complexity had to be minimized not only to enhance the reliability but to facilitate comparison monitoring of the system signals for failure detection. Compensating networks had to be defined which were realistic in terms of maintaining performance in the presence of tolerance effects accompanying the environmental conditions in which the aircraft must operate.

The response amplitudes of the structural modes could be reduced by:

- A reduction in the amount of energy transferred to the structural modes from the input disturbance

- A rapid dissipation of the energy absorbed by the mode.

There are several disadvantages or limiting factors in using a system whose design is based wholly on one or the other of these two fundamental approaches. It was concluded that a system should be designed to make use of the features of both approaches. Considering the first approach, the only means for reducing the amount of energy being transferred is to apply a cancelling signal from some other source of energy (e. g. , a control surface deflection). However, this approach relies heavily upon an accurate knowledge of vehicle dynamics. This is undesirable because of the uncertainty with which the vehicle dynamics are known. Then, too, a system based on cancellation would more than likely require complex scheduling of its parameters because of its sensitivity to variations in vehicle dynamics. Considering the second approach, dissipation of the energy, once it has been absorbed by the mode, can be achieved by augmenting the stability of the modes. Design of a system to provide stability augmentation tends to assure maximum tolerance to variations in vehicle dynamics, which is most desirable. But experience has shown that when two (or more) structural modes are close together in frequency it is difficult (because of coupling effects) to achieve a sufficient increase in the stability to significantly reduce the response amplitudes of the structural modes. A compromise between these two approaches can result in the simultaneous reduction of the structural mode response amplitudes at the sensitive fuselage stations and retention of adequate tolerance to uncertainties or changes in the vehicle dynamics.

To summarize at this point, it is evident that the following design objectives should be used to form the basis for the development of an effective elastic mode suppression system:

- Reduce the response amplitudes of the significant structural modes (generally the lower frequency modes) in terms of the rms acceleration level at all sensitive fuselage stations where ride improvement is a requirement

- Provide maximum tolerance to unpredictable variations in vehicle-system dynamics

- Maintain adequate rigid body dynamics in terms of good handling qualities and in terms of preventing amplified turbulence response at the short-period frequencies

- Minimize unfavorable coupling with the structural modes above the frequency range of the significant structural modes

- Establish an analytical design which permits an economical and reliable hardware implementation.

A necessary condition for suppression of structural flexure is the ability to sense the deflections of the significant modes and to apply control forces to these modes. The deflections of the modes can be measured by conventional sensors (such as accelerometers) suitably located on the vehicle. A corrective control force can be applied through the deflection of one or more of the aerodynamic control surfaces. Hence, a design approach was assumed which would concentrate on the development of a system using conventional sensors to measure the modal deflections and to command corrective action through the aerodynamic control surfaces.

The ride quality can be improved by the elastic mode suppression system (EMSS) independent of the rigid body stability augmentation system (herein referred to as the rigid body SAS). The EMSS functions independently of the SAS to suppress the accelerations at the sensitive fuselage stations due to structural vibrations. To this end, both the rigid body SAS and the autopilot are designed to provide absolute gain stability for all bending modes through the use of low-pass filtering. Conversely, the elastic mode suppression system is designed to have little or no effect over the rigid body frequencies. This approach, however, does not exclude the requirement for a coupled analysis of both the rigid body SAS and the EMSS. There are several advantages to be realized by this separation of system functions:

- This approach permits isolation of problem areas and facilitates preliminary selection of system parameters for both the rigid body SAS and the EMSS.

- Greater flexibility is permitted in the design of both systems. Filter design, sensor placement, and sensor type need not be compromised between rigid body requirements and mode suppression requirements. Sensors for the rigid body SAS can be located to minimize the structural mode components in their output signals, while EMSS sensors can be combined to minimize pickup of rigid body motion. This approach provides a natural decoupling of the two systems. Further, control surfaces to be used in either system can be selected on the basis of their effectiveness for their particular function. Investigations have shown that the surfaces most suitable for control of the short-period mode are often not the most suitable for control of the structural modes and vice versa.

- Impact of equipment failures is greatly reduced and efficient utilization of equipment redundancy according to the value of the function can be realized. For example, the rigid body SAS is considered essential for good control, thus often justifying a triple redundant system for fail operational performance. The mode suppression system is not considered as essential, and thus usually requires only a dual redundant mechanization for failsafe operation.

The variations in vehicle mass, mass distribution, aerodynamics, and atmospheric turbulence require a highly tolerant system to provide satisfactory performance over the entire flight regime. The tolerance of the system depends on the nature of the design. However, it is evident that system complexity can be minimized if active augmentation is applied only over a limited portion of the flight regime. By avoiding an active control attempt over the entire flight regime, the adaptation problem is minimized. Thus, a real need for an active controller at a given flight condition must exist before the added complexity and associated costs can be justified.

In the preceding paragraphs the design objectives and design approach for the elastic mode suppression system were established. This information is applied in the following paragraphs to derive the formulas and constraints used in the synthesis technique.

DEVELOPMENT OF THE SYNTHESIS TECHNIQUE

In general, the synthesis technique used to design an EMSS should have the following properties:

- Facilitate design studies using a very high order mathematical model of the vehicle.

- Enable use of this model in a frequency response form, thereby facilitating incorporation of non-steady aerodynamic effects.

- Capability to compute required controller characteristics directly from specification of meaningful design criteria.

- Permit rapid and accurate evaluation of non-ideal controller characteristics.

The synthesis technique described on the following pages is an attempt to provide a design procedure with these properties.

The technique consists of (1) the specification of the control system filter, and (2) the specification of the sensor configuration. The basic formula used in the synthesis technique specifies the control system filter (i. e. , compensation) requirements in terms of a given sensor configuration and an allocated attenuation requirement on the rms acceleration level at a sensitive fuselage station (e. g. , the pilot's station). Additional requirements are specified in terms of stability constraints and in terms of restrictions resulting from the use of practical (in terms of mechanization) linear filters. These additional requirements enable the designer to specify a sensor configuration. As a consequence, once a reduction in the rms acceleration level (a measure of the ride quality) has been established at the sensitive fuselage stations, it is possible to specify a sensor-filter combination. The synthesis procedure requires a description of the vehicle dynamics in a frequency response format relating output variables of modal accelerations to input variables of wind gust and control surface deflection. Specification of data in this form facilitates incorporation of unsteady aerodynamics, a factor often not included in existing techniques.

For the sake of clarity, it is assumed in the discussion that the first three structural modes are the significant modes in terms of ride quality and, hence, are to be controlled by the EMSS. Actually, any number of modes may be used. But it should help the reader in following the development of the synthesis technique and in the subsequent example to talk in terms of a specific number of structural modes instead of, for example, "n" significant modes.

The approach to be taken toward performance specification for purposes of controller synthesis is to specify the desired reduction in acceleration at each of the resonant peaks of the first three modes. The frequencies of these peaks will be determined from a frequency response plot between normal accelerations and wind velocity, thereby including all known coupling effects. The required percentage reduction in amplitude at each frequency will be estimated from power spectral density plots of acceleration under turbulence at sensitive vehicle stations. Knowing the rms acceleration and identifying its source in terms of individual modes, a design goal for reduction of resonant amplitude can be allocated to each mode. An important part of this allocation is definition of modes which do not cause significant acceleration contribution, so that the required control action at those particular frequencies can be relaxed. The objective of this process is to produce an appropriate allocation of control effort.

With this information, the controller can be specified (as will be shown) and then evaluated at other sensitive stations, and stability margins can be established. If all constraints are not met, the allocation of required reductions in amplitudes at the resonant modes can be modified and the process repeated until all the constraints are satisfied. This results in an iterative process but with an orderly and systematic means of attack. The relating of peak accelerations to rms accelerations and the specification of reductions required in terms of acceleration at a sensitive station are useful methods for expediting the control synthesis process. The final measures of control effectiveness are the actual rms accelerations at the various vehicle stations.

To establish the basic formulas, consider a general block diagram of the control system-vehicle dynamics as shown in Figure 1. This diagram illustrates the relationship (in transfer function form) between the acceleration level measured at a sensitive fuselage station, $\ddot{Z}_c$, and the wind turbulence input, W_g. Clearly, the diagram could be expanded to include all the sensitive fuselage stations, but this is not necessary for the following discussion. The term $\ddot{Z}_s$ is the quantity to be measured by a sensor located at some vehicle station. The quantity $K_cG(S)$ is the filter to be designed for the EMSS. Thus, it is the task of the synthesis process to determine what quantity will be sensed ($\ddot{Z}_s$) and to define a filter $K_cG(S)$ that will result in a specified reduction in the acceleration $\ddot{Z}_c$ in response to the turbulence input, W_g. It is assumed that the aircraft model is known in a frequency response form relating output variables of acceleration to input variables of wind turbulence, W_g, and control surface deflections, δ. Actuation system dynamics can either be included in the aircraft model or be left to be specified as part of the filter, $K_cG(S)$. For the discussion it has been assumed that a satisfactory rigid body SAS has been defined with appropriate filtering to adequately attenuate the stability augmentation signals at the structural mode frequencies.

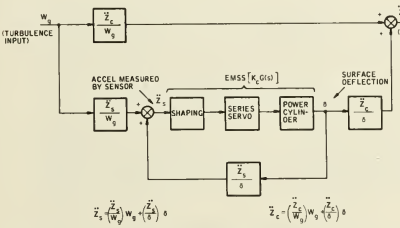


Figure 1. General Block Diagram of Vehicle-System Model

With the design goals in terms of a reduction in acceleration at a sensitive station with augmentation to that without augmentation, one can derive an expression for the required filter characteristic, $K_cG(S)$. Let Y equal the ratio of acceleration at a sensitive fuselage station with augmentation to that without augmentation. Now the acceleration occurring at the sensitive station, Z_c, without augmentation is simply:

$$\ddot{Z}_c\Big)_{\substack{\text{no}\\ \text{Aug}}} = \left(\frac{\ddot{Z}_c}{W_g}\right) W_g \qquad (1)$$

The acceleration at the sensitive station with augmentation is:

$$\ddot{Z}_c\Big)_{\substack{\text{with}\\ \text{Aug}}} = \left[\left(\frac{\ddot{Z}_c}{W_g}\right) + \left(\frac{\ddot{Z}_s}{W_g}\right) \times \left(\frac{K_cG(S)}{1 - \left(\frac{\ddot{Z}_s}{\delta}\right)K_cG(S)}\right)\left(\frac{\ddot{Z}_c}{\delta}\right) W_g\right] \qquad (2)$$

Thus, by the definition of Y we obtain:

$$Y = \left[\left(\frac{\ddot{Z}_s}{W_g}\right)\left(\frac{K_cG(S)}{1 - \left(\frac{\ddot{Z}_s}{\delta}\right)K_cG(S)}\right)\frac{\left(\frac{\ddot{Z}_c}{\delta}\right)}{\left(\frac{\ddot{Z}_c}{W_g}\right)} + 1\right] \qquad (3)$$

This definition of Y implies that Y can be a complex number. The designer is free to pick any Y, real or complex, within certain limits. Clearly, if the control is to reduce the amplitude of the acceleration at the discrete frequency points, then the absolute magnitude of Y chosen must be less than one (i.e., $0 \le |Y| \le 1$). Selecting real values for Y will result in lower gain values for the filter $K_cG(S)$ than will the selection of complex values. Low gain values for the filter are desirable from the standpoint of minimizing the coupling with the rigid body mode and the high frequency structural modes.

We can solve Equation (3) for $K_cG(S)$ to yield

$$K_cG(S) = \frac{1}{\left(\frac{\ddot{Z}_s}{\delta}\right)}\left[\frac{1}{1 - \frac{1}{1-Y}\left(\frac{\ddot{Z}_c/\delta}{\ddot{Z}_c/W_g}\right)\left(\frac{\ddot{Z}_s/W_g}{\ddot{Z}_s/\delta}\right)}\right] \qquad (4)$$

This equation defines the filter characteristics $K_cG(S)$ in a frequency response form. It implies that if one picks the quantity to be measured, $\ddot{Z}_s$, and a value of Y at discrete frequency points (e.g., Y_1, Y_2, Y_3 corresponding to the first three structural mode resonant frequencies), then a gain and phase requirement can be computed for $K_cG(S)$ at these discrete frequencies. If a filter can be designed to exhibit these gain and phase characteristics, then the acceleration $\ddot{Z}_c$ will be attenuated by the values of Y chosen at these frequency points. It is sufficient to specify a filter requirement at only the resonant frequencies. Thus, the resulting filter will not resemble the inverse of the vehicle transfer function as would probably be the case if the accelerations were to be reduced by only a cancellation technique.

The remaining task is to establish a procedure for defining the quantity to be measured, $\ddot{Z}_s$. The objective is to specify the relative modal components of $\ddot{Z}_s$ and not the type of sensor nor the combination of sensors required to actually provide $\ddot{Z}_s$. The sensor combinations can be established by examining the properties of $\ddot{Z}_s$ and comparing

them with the mode shapes. $\ddot{Z}_s$ is considered to be an acceleration quantity having the following form:

$$\ddot{Z}_s = \ddot{Z}_o + K_{\ddot{\theta}} \ddot{\theta} + \sum_{i=1}^{n} K_i \ddot{\eta}_i \qquad (5)$$

where

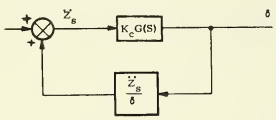
$\ddot{Z}_o$ represents the rigid body plunging mode

$\ddot{\theta}$ represents the rigid body rotation mode

$\ddot{\eta}_i$'s represents the structural modes

In the discussion that follows $\ddot{Z}_s$ is referred to as the sensor complement. Thus, to define $\ddot{Z}_s$ corresponds to defining the relative values of the modal deflections, K_i. To do this we examine the following constraints:

1. A stability constraint is imposed on the system by the stability limits of the feedback loop shown below (see Figure 1):



This loop has the closed loop transfer function given by:

$$H(S) = \frac{K_c G(S)}{1 - (\ddot{Z}_s/\delta)\ K_c G(S)} \qquad (6)$$

H(S) is unstable whenever the roots of its denominator lie in the right-half plane of a root locus diagram. The denominator will be unstable at the frequency at which

$$\phi_{s_\delta} + \phi_g = 0° \qquad (7)$$

and

$$\left| \left(\frac{\ddot{Z}_s}{\delta} \right) K_c G(S) \right| \geq 1 \qquad (8)$$

where

ϕ_{s_δ} = phase angle of the transfer function $(\ddot{Z}_s/\delta)$

ϕ_g = phase angle of the filter $K_c G(S)$

Since one cannot reduce the amplitude of the bending mode accelerations by gain stabilization, it is necessary to provide phase stabilization by proper selection of $\phi_{s_\delta} + \phi_g$. Because of uncertainties in the mathematical model, it is necessary to require that over the range of frequencies where gain stabilization is not attainable nor desirable, the following constraint be applied:

$$90° \leq \phi_{s_\delta} + \phi_g \leq 270° \qquad (9)$$

This constraint assures that the stability of the modes will at least be increased and provide a minimum phase margin of 90 degrees. At frequencies above the modes of interest (i.e., above frequencies where augmentation is to be applied) it is desirable to provide gain stabilization to eliminate this constraint on the phase angle.

With the above constraint on the phase angle of the sensor-filter combination, the problem is to find a solution which, over the frequency range of interest, will simultaneously satisfy this constraint and the phase constraint imposed by the filter equation [Equation (4)]. These two constraints are repeated here for reference:

$$K_c G(S) = \frac{1}{\ddot{Z}_s/\delta} \left[\frac{1}{1 - \frac{1}{1 - Y} \left(\frac{\ddot{Z}_c/\delta}{\ddot{Z}_c/W_g} \right) \left(\frac{\ddot{Z}_s/W_g}{\ddot{Z}_s/\delta} \right)} \right] \qquad (10)$$

$$90° \leq \phi_{s_\delta} + \phi_g \leq 270°$$

To impose the constraints defined by Equation (10) we observe that:

$$\phi_{s_\delta} + \phi_g = \text{phase angle of} \left[\frac{1}{1 - \frac{1}{1 - Y} \left(\frac{\ddot{Z}_c/\delta}{\ddot{Z}_c/W_g} \right) \left(\frac{\ddot{Z}_s/W_g}{\ddot{Z}_s/\delta} \right)} \right] \qquad (11)$$

Defining:

$$N = \frac{|\ddot{Z}_c/\delta|}{|\ddot{Z}_c/W_g|}$$

$$V = \frac{|\ddot{Z}_s/W_g|}{|\ddot{Z}_s/\delta|}$$

ϕ_1 = [phase of $(\ddot{Z}_c/\delta)$ - phase of $(\ddot{Z}_c/W_g)$]

ϕ_2 = [phase of $(\ddot{Z}_s/W_g)$ - phase of $(\ddot{Z}_s/\delta)$]

and restricting $\phi_{s_\delta} + \phi_g$ according to Equation (9), we obtain a minimum attainable value on Y, the attenuation factor.

$$Y \geq 1 - VN \cos(\phi_1 + \phi_2) \qquad (12)$$

At any flight condition, N and ϕ_1 are defined for a given vehicle station. Thus, if we pick a quantity to be measured, $\ddot{Z}_s$, we can determine the lower bound on Y (defined as Y_{min}) in order to assure that the stability constraint (10) is satisfied. That is

$$Y \geq Y_{min} = 1 - VN \cos(\phi_1 + \phi_2) \qquad (13)$$

If we assign a value to Y, say Y_i, at a given frequency (ω_i), then we must find a quantity to be measured, $\ddot{Z}_s$, such that

$$Y_i \geq Y_{min}(\omega_i) \qquad (14)$$

or else we cannot simultaneously satisfy the stability constraint and provide the attenuation $Y = Y_i$. The above condition is a necessary condition for specifying $\ddot{Z}_S$, but it is not a sufficient condition.

2. To further specify the requirements for the quantity to be measured, $\ddot{Z}_S$, it is necessary to examine other design objectives and analyze the practical constraints imposed on the filter, $K_c G(S)$. The following design objective imposes a severe constraint on the type of filters and on $\ddot{Z}_S$.

- Minimize unfavorable coupling with structural modes above the frequency range of the first three modes. To assure that this objective is met we impose the condition

$$\left| \left(\frac{\ddot{Z}_S}{\delta} \right) \left(K_c G(S) \right) \right| < 0.5$$

at modal frequencies above the third mode.

3. Now it is assumed the filter $K_c G(S)$ will be a linear filter made up of first or second order elements and that the poles of the filter will be stable. Thus, the design objective of high frequency gain stabilization, for practical purposes, imposes an additional constraint on $K_c G(S)$. To meet this objective the gain of the filter will be decreasing with increasing frequency at a rate of at least six db per octave over the frequency range of the significant modes (e. g., the first three modes). From this, it becomes evident that if the same amount of attenuation (or more) is required at the second or third mode than as is required at the first mode frequency, then

$$|K_i| \geq \left(\frac{f_i}{f_1} \right) |K_1| \text{ for } i = 2, 3$$

(where K_i is the i^{th} mode deflection) to compensate for the loss in gain due to the filter $K_c G(S)$. For gain stabilization of the fourth and higher modes (assuming three significant modes) it is generally desirable to minimize K_i for $i = 4, 5, \ldots$.

4. A further consideration is the amount of phase or gain change with frequency required of the filter $K_c G(S)$. Sensor configurations should, in general, be chosen to minimize large phase or gain change requirements with frequency for the filter $K_c G(S)$ to facilitate the hardware implementation.

5. It is an objective to minimize the coupling with the rigid body mode. This can be accomplished by placing the sensor in a position where the accelerations due to flexibility are the largest with respect to those due to rigid body motions. The coupling can be further reduced by minimizing the gain of the filter at the rigid body frequencies.

All of the above considerations enable one to essentially define the desired sensor complement, $\ddot{Z}_S$. To summarize, the sensor complement should be chosen to satisfy the following constraints:

- $Y_{min} \leq Y_i$
- $|K_i| \geq \left(\frac{f_i}{f_1} \right) |K_1|, i = 2, 3$ (assuming 3 significant modes)
- Minimize the variation in gain and phase changes with frequency required of the filter $K_c G(S)$
- Maximize the ratio of acceleration due to flexibility (of the significant modes) over the acceleration due to rigid body motions measured by the sensor(s)
- Minimize the gain of the sensor-filter combination at frequencies beyond the significant structural modes.

Once the quantity to be measured, $\ddot{Z}_S$, has been defined, (according to the above rules) realistic sensor locations and combinations of sensors can be examined to find one which will best provide $\ddot{Z}_S$. With the actual sensor configuration defined, the filter requirements can be computed for that sensor configuration through the use of Equation (4). Using these filter requirements, a practical filter can be designed. Finally, with the filter designed, the problem can be worked in reverse to check the "goodness" of the entire system. Thus, compromises in filter design can be analyzed to determine their effect on performance. It must be kept in mind that Equation (4) does not have to be satisfied at all frequencies nor is its solution the only solution to the problem. Once a solution has been obtained, variations can be made on it to yield a possible better solution.

APPLICATION TO A LARGE SST VEHICLE

The synthesis technique is demonstrated through an example using the SST vehicle configuration (Figure 2) as the model vehicle. For the example, only the pitch axis was considered with the assumption of quasi-steady aerodynamics. Four

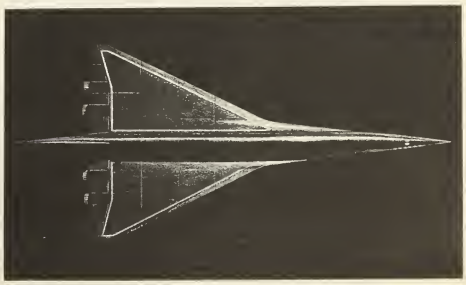


Figure 2. SST Vehicle Configuration

symmetric structural modes were represented, each with an assumed structural damping ratio of 0.015. The following paragraphs describe the design of an EMSS for the SST at the Mach 0.54, 5000-foot flight condition. A heavy configuration (585,000 pounds) was selected. This represented one of the worst conditions in terms of turbulence effects. The system defined at this condition was also evaluated at the Mach 0.9, 30,000-foot flight condition to determine its tolerance.

Before the synthesis technique was applied a qualitative analysis was made to determine the best control surface(s) to use and to establish any constraints on the sensor configuration which would be peculiar to the SST vehicle. There are four separate aerodynamic control surfaces distributed along the trailing edge of each wing (Figure 2). The pair farthest outboard (number four pair) is the most attractive for use in the mode suppression system. Examination of the node lines (ignoring aerodynamics) for the symmetric bending modes shows that the number four pair is the only pair which definitely qualifies for effective force application. A possible alternate for a somewhat restricted portion of the flight regime would be the number three surface pair. Use of the number four pair offered several advantages. These advantages include:

- No direct conflict in authority requirements arise since the outer elevons (number four pair) were employed only during the geardown operating condition. The number three pair is dwelled between Mach 0.95 and 2.4.

- Coupling with the rigid body SAS is minimized due to the reduced surface effectiveness at the rigid body frequencies.

- Use of the outer surface pair has the advantage in that the phase relationship between the generalized forces and the aerodynamic control forces is more consistent (that is, the aerodynamic control forces tend to act only on the control surface itself). It is less affected by unsteady aerodynamic effects and by changes in flight condition because of the relatively short wing chord at the wing tip.

- The outboard surfaces have the least inertia. This results in a lower ratio between inertia forces (due to control surface masses) and the generalized aerodynamic forces, thus the "tail-wag-dog" effects are minimized. This "tail-wag-dog" (TWD) effect is evidenced in the system transfer function by a pair of lightly damped zeros. Because of their low damping ratio and their variation in frequency with surface effectiveness, compensation electronically is difficult. A TWD frequency in excess of the active mode suppression range appears essential if mode suppression is to be performed using aerodynamic surfaces. This frequency is kept high by minimizing the surface inertia.

- Use of the number four surface pair requires the addition of appropriate series servos to actuate these surfaces. This has the advantage in that there would be no conflict with servos presently being used for rigid body stability augmentation. Two disadvantages are the requirements for a high-temperature servo and the added weight penalty.

Use of the inboard surfaces may offer some benefit, particularly in providing some force decoupling among the bending modes. But for this example, only the number four surface pair was considered since it appeared to be the most promising candidate for effective force application.

Linear accelerometers have a significant advantage over rate gyros for use as the bending mode sensors on the SST. Accelerometers have superior high-temperature capabilities, thus facilitating their placement near the outer elevons. Because of the temperature problem, placement of rate sensors would be restricted essentially to the fuselage. This restriction greatly reduces the potential capability of the rate sensor in the mode suppression system, except in a possible combination with an accelerometer. Examination of the mode shapes along the fuselage indicated there was no single rate sensor location which would provide proper phasing for all the first four structural modes. Two or more rate sensors could be blended, but examination of the mode shapes indicates that the resulting signal would be rather sensitive to variations in conditions. Placement of an accelerometer near the outboard surface pair (impractical for a conventional rate sensor because of temperature problems) tends to provide more uniform phasing between the control force applied and the resulting modal deflections. For these reasons, it was concluded that accelerometers offer the greater advantage as a candidate sensor.

The acceleration levels experienced by the free airplane in turbulence can readily be determined from power spectral density (PSD) plots of the mean square acceleration. These plots were obtained using the wind model described in Appendix A. Figure 3 is a PSD plot of the

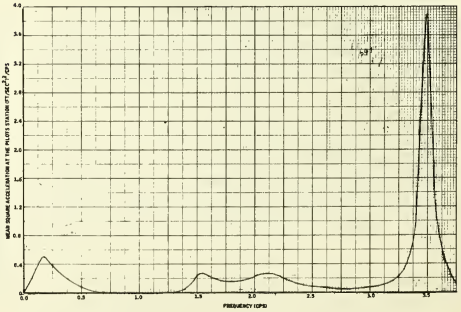


Figure 3. Mean Square Acceleration at the Pilot's Station versus Frequency for a Unit Gust Input. M = 0.54; h = 5000 ft; W = 585,000 lbs

acceleration at the pilot's station at the Mach 0.54, 5000-foot flight condition. The frequency range of the plot covers the significant response amplitudes up through the fourth structural mode. The pilot's station and the aft cabin were found to be the two most sensitive fuselage stations in terms of a degradation in the ride quality due to flexibility on the SST. Calculation of the rms acceleration level from Figure 3 shows it to be on the order of 0.04 g's/ft/sec rms of gust amplitude. Thus, for a nominal gust amplitude of 4 ft/sec, the rms acceleration occurring at the pilot's station would be 0.16 g's. Figure 4 is a plot which indicates the acceleration level as a function of frequency which can be comfortably tolerated by a human being. This plot (see Reference 5) is a composite made by assimilating the data from References 1, 2, and 3. Although the duration of the turbulence has not been taken into account, this plot indicates that an rms acceleration level on the order of 0.16 g's is large in terms of passenger comfort. Sufficient data was not available to establish a specific design objective in terms of the amount of reduction required in acceleration levels at the sensitive stations over the flight envelope. However, the order of magnitudes presented above indicate that a design objective of a 50 percent reduction in the acceleration levels due to structural mode responses is a reasonable goal.

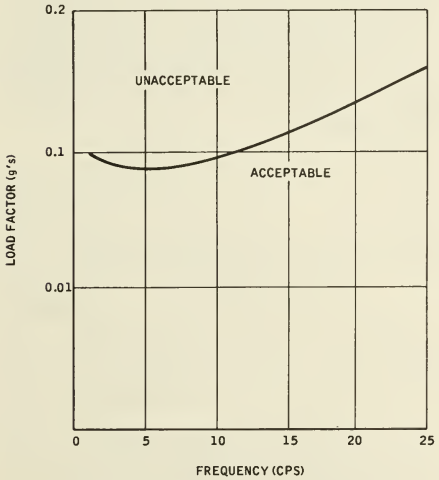


Figure 4. Typical Human Response to Vibration Characteristics

For the example, the following values of Y were assumed:

	bending frequencies	attenuation factor
at	f = 1.55 cps	$Y_1 = 0.5$
	f = 2.15 cps	$Y_2 = 0.5$
	f = 3.5 cps	$Y_3 = 0.25$

(15)

Application of the synthesis technique established the following set of requirements for the sensor complement:

a. $\ddot{\bar{Z}}_o$ term be minimized

b. $K_{\theta}^{\ddot{\bullet}}$ be minimized

c. $K_2/K_1 \geq 1.4$

d. $K_3 < 0$ and $|K_3| \geq 2\ |K_1|$

e. $K_4 > 0$

A sensor complement was selected which tends to satisfy these requirements. It was as follows:

$$\ddot{z}_s = (\ddot{z}_{26} - \ddot{z}_{18}) = -477.2\ \ddot{\theta}_o + 1.479\ \ddot{\eta}_1 + 3.34\ \ddot{\eta}_2 - 3.625\ \ddot{\eta}_3 + 0.651\ \ddot{\eta}_4 \quad (16)$$

This sensor complement was made up of three accelerometers. An accelerometer was placed at the two wing tips and their signals were averaged together to provide the net acceleration, $\ddot{z}_{26}$. A third accelerometer was placed near the aft end of the fuselage ($\ddot{z}_{18}$). With this sensor complement and the values of Y given by (15) the following filter requirements were specified by Equation (4):

$$\text{at } f = 1.55 \text{ cps} \quad |K_c G(S)| = 0.252 \times 10^{-4} \frac{\text{rad}}{\text{ft/sec}^2}, \phi_g = 214^\circ$$

$$f = 2.15 \text{ cps} \quad |K_c G(S)| = 0.199 \times 10^{-4} \frac{\text{rad}}{\text{ft/sec}^2}, \phi_g = 117^\circ \quad (17)$$

$$f = 3.5 \text{ cps} \quad |K_c G(S)| = 0.425 \times 10^{-4} \frac{\text{rad}}{\text{ft/sec}^2}, \phi_g = 121^\circ$$

The following filter was designed to meet these requirements:

$$K_c G(S) = 10^{-3} \left[\frac{S^2 + 2(0.07)\ 10S + 10^2}{S^2 + 2(0.7)\ 10S + 10^2}\right] \left[\frac{S^2 + 2(0.2)\ 30S + 30^2}{S^2 + 2(0.2)\ 5S + 5^2}\right] \left(\frac{5}{30}\right)^2 \left[\frac{60^2}{S^2 + 2(0.5)\ 60S + 60^2}\right] \quad (18)$$

A frequency response of this filter is shown in Figure 5. The 60 rad/sec lag was used to represent the power cylinder dynamics. No special

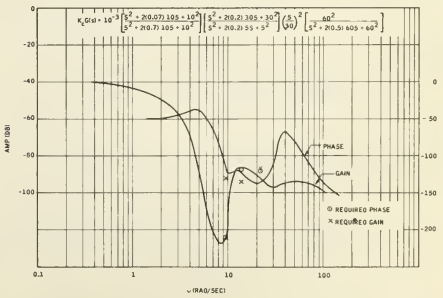


Figure 5. Frequency Response of $K_c G(S)$

effort was made to define the simplest and most tolerant filter since its purpose was solely to demonstrate the validity of the synthesis technique. Figure 6 is a frequency response of the vehicle-system combination. The following points should be noted about this diagram.

- Since $90° \leq \phi_{s_\delta} + \phi_g \leq 270°$ at essentially all points between the first and fourth modes, this phase constraint has been satisfied.
- The fourth mode is gain stable with a gain margin of 5 db.
- The system gain margin is 10 db.

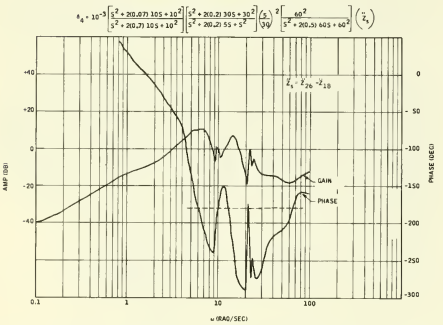


Figure 6. Open Loop Frequency Response of Vehicle-Controller Dynamics

Figures 7 and 8 show the PSD plots of the acceleration at the pilot's station and at the aft cabin with this control system at the Mach 0.54, 5,000-foot flight condition. It is observed that the required attentuation has been met at the first and second modes. It falls short of the objective of 0.25 at the third mode since the filter provided only half the required gain as evidenced in Figure 5. Nevertheless, the constraints of the problem

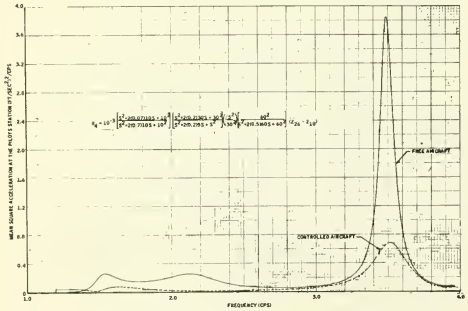


Figure 7. Mean Square Acceleration at the Pilot's Station versus Frequency for a Unit Gust Input. M = 0.54; h = 5000 ft.; W = 585,000 lbs.

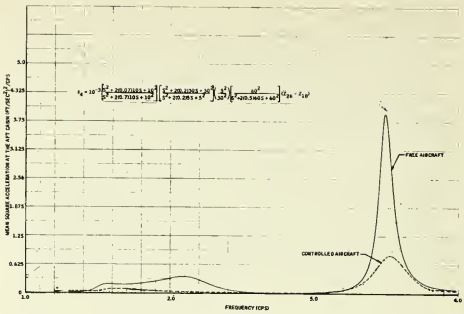


Figure 8. Mean Square Acceleration at the Aft Cabin versus Frequency for a Unit Gust Input. M = 0.54; h = 5000 ft.; W = 85,000 lbs.

have essentially been satisfied. The percent reduction in the rms acceleration at the pilot's station was 42 percent and at the aft cabin it was 41 percent. Thus, the rms acceleration levels have been favorably reduced while at the same time adequate stability margins have been maintained. This same system was also evaluated at the Mach 0.9, 30,000-foot flight condition. Figures 9 and 10 illustrate the PSD plots of the acceleration at the pilot's station and at the aft cabin. Essentially the same performance was obtained at this condition as was obtained at the Mach 0.54, 5,000-foot flight condition.

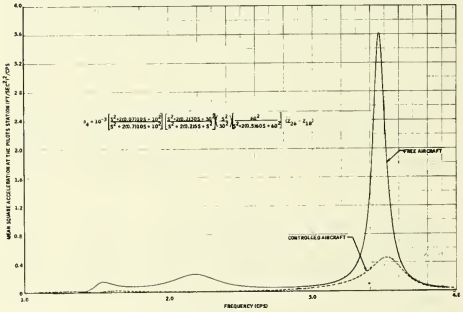


Figure 9. Mean Square Acceleration at the Pilot's Station versus Frequency for a Unit Gust Input. M = 0.9; h = 30,000 ft.; W = 585,000 lbs.

It was found that with this system some coupling with the rigid body mode was evident. To avoid this coupling, the filter would have to be modified slightly to provide additional attenuation at the rigid body mode. Since the objective herein was to demonstrate the synthesis technique primarily over the frequency range of the structural modes, this coupling was ignored. Of course, in any real application, this coupling would be taken into account in the design of the filter.

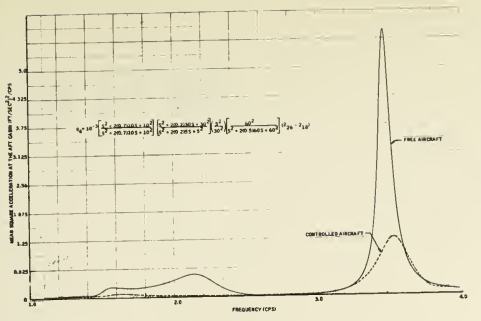


Figure 10. Mean Square Acceleration at the Aft Cabin versus Frequency for a Unit Gust Input. M = 0.9; h = 30,000 ft.; W = 585,000 lbs.

CONCLUSIONS

The example described in the paper illustrated that the synthesis technique is an effective tool for use in the design of an elastic mode suppression system. An understanding of a few simple formulas and a knowledge of frequency response analysis enables the control system designer to readily acquire a working knowledge of the technique. The synthesis procedure has the following advantages.

- It provides insight into the nature of the problem through the use of PSD plots and vector diagrams. This feature is especially apparent when analyzing the inter-modal coupling.
- It provides the capability to handle unsteady aerodynamics due to the reliance upon a frequency response format for the vehicle data.
- Once a filter-sensor combination has been selected the synthesis procedure can be worked in reverse to establish the resulting performance. This feature enables the designer to readily assess any compromises in system design.
- Given the vehicle data in frequency response form all the computations can be easily worked by hand. The process, however, could be programmed on a computer should it be desirable to do so.
- The synthesis technique is not limited by the order of the vehicle dynamics.

In considering the final point above, it should be remembered that while the technique can be applied to a vehicle, regardless of the number of structural modes, there are a few applications where the technique would be of limited use. For example, it is desirable to separate the rigid body control problem from the structural mode control problem. If this is not possible through the selection of the control forcing function, sensor configuration and filtering, it may be difficult to simultaneously provide both rigid-body handling qualities and adequate structural mode suppression. In one respect, optimal control techniques may provide an advantage over the technique described herein. Use of optimal control techniques may offer an advantage in designing a system where several forcing functions are considered simultaneously. Clearly the technique described would become rather cumbersome in this case.

It is anticipated this synthesis technique would be applicable to a variety of flexible aircraft but in particular to large transports such as the SST. Further development of this technique is expected to lead to application in other areas such as stress reduction in vehicles like the Boeing 747 and the Lockheed C-5A.

APPENDIX A
WIND MODEL

The turbulence power spectral density function used in the analysis is given by:

$$\Phi(\Omega) = \sigma_G^2 \frac{L}{\pi}\left[\frac{1+\frac{8}{3}(1.339L\Omega)^2}{\left[1+(1.339L\Omega)^2\right]^{11/6}}\right]\frac{(\text{ft/sec})^2}{\text{rad/ft}} \qquad \text{(A1)}$$

where

$\Omega = \frac{\omega}{V}$ = reduced frequency (rad/ft)

V = aircraft velocity (ft/sec)

L = characteristic length (ft)

σ_G = rms gust velocity (ft/sec)

and

$$\int_0^\infty \phi(\Omega)\, d\Omega = \sigma_G^2$$

REFERENCES

1. Lippert, S., "Human Response to Vertical Vibration", Douglas Aircraft Co., Report SM-20021, 18 October 1945.

2. Parks, D. L. and Beaupeurt, J. E., "Human Reaction to Low Frequency Vibration", The Boeing Company, Wichita Division, Report No. D3-3512, 24 July 1961.

3. Parks, D. L., "A Comparison of Sinusoidal and Random Vibration Effects on Human Performance", the Boeing Company, Wichita Division Report No. D3-3512-2, 28 July 1961.

4. Notess, Charles B., "A Triangle-Flexible Airplanes, Gusts, Crew", Cornell Aeronautical Laboratory Inc., Full-Scale Division Memorandum No. 343, May, 1963.

5. Wykes, John H. and Mori, Alva S., "An Analysis of Flexible Aircraft Structural Mode Control, Part I", North American Aviation,

Inc. / Los Angeles Division, Technical Documentary Report No. FDL-TDR-65, August 1965.

ACKNOWLEDGMENTS

The author wishes to acknowledge the Lockheed-California Company for providing the vehicle data and technical assistance for this paper. In particular, I would like to thank Messrs. E. O. Throndsen, J. A. Flapper, and H. P. Weinberger for their very significant contributions.

I wish to also acknowledge Messrs. J. C. Larson and R. C. Hendrick of Honeywell for their contributions through their leadership and technical assistance.

No. 67-572

WHERE ARE WE NOW WITH ALL-WEATHER LANDING?

by

J. A. GORHAM, H. K. RICHTER,
and
F. B. LEE
Lockheed-California Company
Burbank, California

AIAA Paper
No. 67-572

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.09, 7.11, 8.08

-- NOTES --

WHERE ARE WE NOW WITH ALL-WEATHER LANDING?

J. A. Gorham, Deputy Chief, CL-1011 Avionics and Controls
H. K. Richter, Avionics & Controls Division Engineer
F. B. Lee, CL-1011 Flight Controls

Lockheed-California Company
Burbank, California

ABSTRACT

Recently there has been an upsurge of activity directed towards a wide scale implementation of lower weather operating limits for commercial and military transport aircraft. This activity follows many years of in-service operation to 200' and 1/2 mile limits and concurrently, study and experimentation aimed towards an ultimate all-weather operational capability. The evolution of the systems concepts which are being employed in present generation aircraft is described together with the constraints and incentives which have guided the development. These concepts are examined in terms of the operational techniques and design criteria needed to meet the required performance and safety standards. The likelihood of these concepts remaining viable down to an ultimate zero visibility operation is examined in light of present experience and the problem areas are discussed. In particular, the overall systems aspects which include the role of the pilot and the influence of environmental variables are considered. Some of the known problems posed by the present approach to all-weather operation are examined and some suggestions are offered for a reorientation of present effort on the basis of a systems approach towards achievement of all-weather landing.

Evolution of Airline Instrument Approach Systems

Ever since the memorable day of September 4, 1929, when James Doolittle made the first blind landing and takeoff entirely by instruments, there has been guidance and control system design activity of varying magnitude and intensity directed toward the realization of the dream of regular all-weather operation of commercial transport airplanes. This requirement is the kind that inherently poses a technical challenge to many people. In particular, research and development organizations all over the world have devoted much time and energy to developing systems which could make the objective come true. A number of military aircraft, both fighter and transport have been successfully fitted with instrument landing systems which can operate under very low visibility conditions. Aircraft carriers have probably had the most urgent need for such systems and "all-weather" operations are now almost routine using highly developed instrument and radar guidance techniques.

Progress in commercial aviation, however, has been different for many reasons. The systems equipment content of a commercial airplane is less sophisticated and less advanced than its military counterpart. However, there has been greater progress in commercial aviation toward the development of simpler but highly reliable and usable navigation and communications systems. The high utilization of commercial airplanes in congested airways and terminal areas has been largely responsible for this trend. Certainly the airlines make more regular use of instrument landing facilities at the airports at which they operate than military transport airplanes. Additionally ILS installations have been standard at major commercial airfields for many years. Prior to 1949, commercial aviation made use of ILS installations by utilizing approach guidance systems comprising deviation indicators displaying errors from glide slope and localizer beams. The crew were able to fly the airplane manually by zeroing the displayed error and, in many cases, airplanes were able to operate down to limits of 200 feet and half a mile visibility.

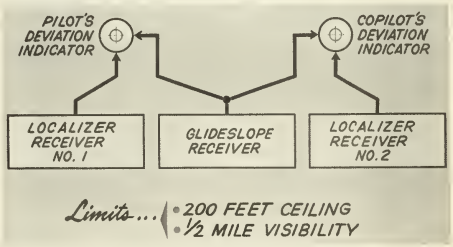


Fig. 1 Standard Approach Guidance Before 1949

In 1949, Sperry introduced the zero reader into airplane service and the approach system became more sophisticated. A computer in the airplane could now equate information from the airplanes attitude against the deviation from the glide slope and localizer beams and thus provide to the pilot a more sophisticated command indication displayed on the zero reader indicator. This system required the installation of more equipment in the airplane but resulted in improved stability of the airplane on the ILS beams and a better consistency

67-572

of performance. This resulted in fewer go-arounds as a result of missed approaches. The system was used successfully for many years and became an accepted means of carrying out instrument landings.

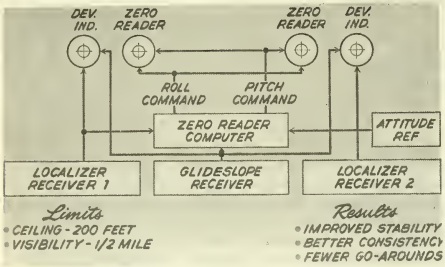


Fig 2 1949 Sperry Introduces the Zero Reader

During the 1950's airplanes began to be equipped with an autopilot, previously used as a cruise aid device, which included an approach coupler. The approach coupler derived its basic guidance information from the glide slope and localizer errors and also from other sensors such as attitude and heading. By this means automatic control of the airplane on the ILS beams became feasible. At about the same time, the concept of integrating the deviation indicator, the zero reader indicator, the artifical horizon and compass information into two basic centrally situated instruments was evolved and implemented.

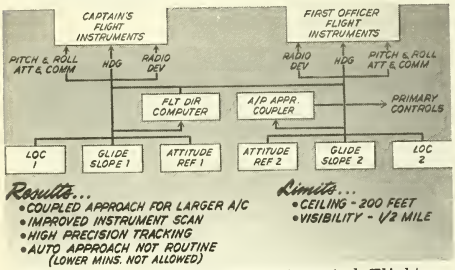


Fig. 3 Autopilot Coupler & Integrated Flight Instruments Introduced

This development enabled flight director manual instrument approaches to be conducted with the advantage that the pilot could survey the command and the resulting flight attitudes on two basic flight instruments - the Horizon Director Indicator and the Course Deviation Indicator. Thus the capability of carrying out approaches to an airfield equipped with ILS, either on the integrated flight instrument system or by the automatic pilot, became a reality. The results of this were to improve the quality of coupled approaches for large airplanes by providing reduced instrument scanning for the crew and a higher precision of tracking of the ILS beams. Despite this increase in sophistication, the weather limits still remained at a ceiling of 200' and a visibility of half a mile.

Shortly after the introduction of autopilot approach couplers, a general unreliability of the total system in the coupled approach mode created an atmosphere of distrust by operating personnel which has persisted until quite recently. The problems which arose were due partly to the general unreliability of the avionic equipment associated with the automatic approach coupler, the vagaries of the ILS beams, the inflexibility of the capture modes and also unfamiliarity with the system by maintenance personnel and the flight crews who used it. At various stages, autopilots were distrusted so much that some airlines either prohibited their use or removed the equipment entirely. However, as designs improved, pilots began to appreciate the work saving feature of the automatic coupled approach. This factor accompanied by firm policy pressures by their management caused a trend towards more regular use of the equipment. Despite this, however, the automatic approach coupler of todays' airplanes still does not possess the degree of reliability and consistency of performance required by many pilots before they would use automatic approaches below the present 200' limit with confidence. There are of course, exceptions to this rule and there are some airlines and certain systems which give excellent and consistent performance.

Concurrent with the airlines experience of increased sophistication and capability of airborne equipment for coupled approaches, many experimental establishments throughout the world were conducting tests of similar equipment modified so as to be able to continue the approach through to the landing in the fully automatic mode. Probably outstanding in this field of activity, is the Blind Landing Experimental Unit (BLEU) of the Royal Aircraft Establishment in England, who initiated a comprehensive program of research and development in 1948. This program was aimed at solving many of the problems involved in producing a viable automatic landing system for transport airplanes based on current autopilot designs and conventional ILS beams. To date, well over 15,000 experimental landings in several types of military and commercial transports have been completed in all types of weather and BLEU has accumulated a body of knowledge and experience in all-weather landing which probably matches that of any other group in the world. In the United States, the FAA has been conducting many experimental programs all leading toward lower visibility landings for commercial aviation. In 1962 they carried out a development program of over 1,000 landings in a DC-7 using the BLEU system concept and are currently conducting all-weather landing trials in a Convair 880 airplane. In addition to these two programs, they have many other evaluations of peripheral equipments which will contribute to all-weather operations. In France, more than 5,000 automatic landings have been completed successfully since 1962 in Caravelles equipped with an all-weather system developed by an American

manufacturer. It is a sobering thought, however, that the overall total of some 25,000 experimental landings carried out over the past 15 years would be exceeded by the airlines within a few months of regular use of automatic landings.

Where Are We Now

Since the memorable day 38 years ago when the first blind landing was completed, much technical progress has been made. Many thousands of successful automatic landings have been completed on an experimental basis in all types of weather conditions and at many different airfields. Airlines all over the world are commencing to purchase equipment or equipment additions which will enable them to operate to weather minima lower than the present 200' and half a mile. In 1963, a ten-day symposium sponsored by IATA was held in Lucerne, Switzerland to survey the general situation and progress of the airlines toward blind landing. The objective was to formulate ground-rules so that further development would be conducted internationally on a compatible basis, with particular emphasis on ground aids. ILS beam characteristics were discussed and an initial definition of required performance to lower altitudes was made. Weather minima categories were defined, enabling operation to eventual touchdown to be made on a progressive basis under specified weather conditions. These weather minima categories were as follows:

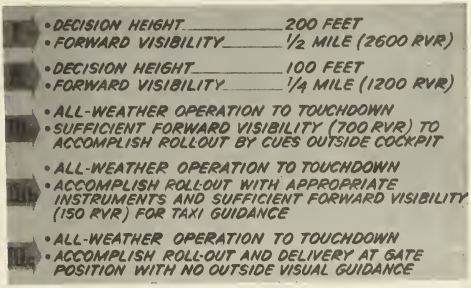


Fig. 4 Weather Minima Categories

Many airlines have now purchased and fitted equipment with a capability of operating to Category II weather minima, although no airline is as yet operating to the Category II minimum RVR of 1200'. Several airlines have permission, however, to operate to 1600' RVR as a progressive step towards Category II minima. In order to define the standards of performance required for Category II systems, the Federal Aviation Agency has issued several comprehensive advisory circulars in addition to the basic regulations concerning the individual use of the various airborne systems involved. Fig. 5 shows broadly the type of accuracy criteria which is specified by the FAA.

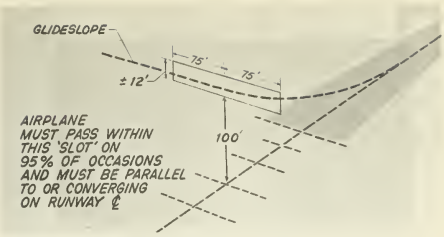


Fig. 5 Accuracy Criteria at 100' for Category 2 Approaches

The applicant must demonstrate by actual flight test that an airplane fitted with a given configuration of equipment will arrive at the 100' position point within a "slot" 150' wide and 24' deep. The airplane must pass within this "slot" on 95% of the occasions during equipment performance demonstrations and must be either parallel to or converging on the runway centerline. For operational demonstration, 85% of the approaches attempted must pass through this "slot." The equipment content itself, naturally involves a number of alternative configurations which the FAA define are a minimum to carry out the operation since various airlines have preferences for different types and designs of airborne equipment.

Figure 6 shows in chart form, typical combinations of equipment which will meet the requirement of the FAA Advisory Circular, AC120-20, defining Category II airborne equipment requirements. It can be seen that basically, the airplane must possess ILS localizer and glide slope receivers. Flight control guidance in the form of an approach coupler (either all axis or split axis - roll or pitch) and a flight director system (either single or dual) may be used. Identification of decision height must be provided in the form of a radio altimeter, a barometric altimeter of demonstrated accuracy or a marker beacon receiver which will define the height because of its

FAA ADVISORY CIRCULAR AC NO. 120-20

ITEM	ALTERNATE CONFIG. ➡	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	ILS LOCALIZER RECEIVER	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	ILS GLIDESLOPE RECEIVER	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
FLIGHT CONTROL GUIDANCE																			
3	AUTOMATIC APPROACH COUPLER ALL AXES	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	AUTOMATIC APPROACH COUPLER SPLIT AXIS	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0
5	FLIGHT DIRECTOR	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
IDENTIFICATION OF DECISION HEIGHT																			
6	RADIO ALTIMETER	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0
7	BAROMETRIC ALTIMETER **	0	0	2	2	0	0	0	0	2	2	0	0	0	0	2	2	0	0
8	MARKER BEACON RECEIVER	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1
MISSED APPROACH ATTITUDE GUIDANCE																			
9	ATTITUDE GYRO INDICATOR WITH CALIBRATED PITCH ATTITUDE MARKINGS	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0
10	FLIGHT DIRECTOR PITCH COMMAND	0	1	1	0	0	1	0	1	1	0	0	1	0	1	1	0	0	1
11	COMPUTED PITCH COMMAND	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0
12	AUTOTHROTTLE	0	0	0	0	0	0	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
13	INSTRUMENT FAIL WARNING	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	RAIN REMOVAL SYSTEM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

* REQUIRED UNLESS PILOT WORKLOAD SHOWN NOT INCREASED OVER WORKLOAD WITH AUTOMATIC APPROACH COUPLER (ALL AXES)

** = ACCURACY MUST BE DEMONSTRATED

Fig. 6 Category II Operations by All Jet A/C Except 4-Engine Jet

67-572

fixed position from the runway threshold when the aircraft passes over it. To permit a safe go-around procedure following a missed approach, the airplane must be equipped with either a calibrated attitude gyro indicator, a flight director pitch command, or a computed pitch command. Other items of equipment which may be required are an automatic throttle for control of airspeed, an instrument failure warning system to identify disparities between individual elements of dual system installations and a rain removal system. It can be seen from the chart that the basic equipment required for Category II operation is familiar in many basic respects to equipments used years ago for simple coupled approaches. Economic pressures and the reluctance to depart from familiar equipment and operating techniques have guided the development of all-weather landing systems on an evolutionary basis and this is the situation which exists today.

During the 1963 symposium in Lucerne, discussion on the floor indicated that there was a difference in the overall systems approach to all-weather landing between authorities in the United States and Britain. The position in the United States was stated as requiring an operational system designed to allow pilot intervention in the approach and landing if required. Use of either partial or full automatics was envisaged at that time. The British approach was a resolute one - that all-weather landing would only be accomplished by a fully capable automatic system with the pilot in a supervisory or monitoring role only. Perhaps this difference of opinion has also been responsible in part for the delay in a wide spread and rapid development of systems to lower the present weather limits due to the economic implications. Independent development of a complex avionics system to satisfy the requirements of automatic landing would necessitate a financial commitment sufficiently large to ensure that most companies would only proceed when the requirement becomes firm enough to support a market sufficiently large to ensure a reasonable return on investment.

During 1966, as a result of development programs conducted by various aircraft and instrument manufacturers, the airlines started to purchase equipment capable of automatic touchdown under Category II weather conditions. The principle of operation was basically that which was developed by BLEU comprising a relatively standard automatic approach using ILS to approximately 50-70' and a radio altimeter for flare control to the runway surface. At about the same time, the Airline Pilot's Association (ALPA) and the FAA formulated a firmer policy that automatics were the best evolutionary route to landing in very low visibilities. The pilot's role in the operation, however, is still as yet unresolved. The standards for automatic touchdown in terms of performance accuracy and the effect of failures are in the process of being defined by the FAA and it is expected that a circular defining these requirements will be issued shortly. So it appears that we are currently at the point of breaking through the 200' and half a mile visibility limits which have existed for the past 20 years.

System development has also been constrained, paradoxically, by the airlines and the FAA themselves. Traffic growth in commercial aviation has been so rapid and extensive and the need for new airports so great that there has been an understandable reluctance to install new ground based guidance equipment which might improve the potential performance of all-weather landing systems. In addition, because of the economic pressures to operate airplanes at the maximum utilization rate, there has been a reluctance to introduce any radically new airborne equipment designs until a reasonable reliability level has been established. Understandably, the airlines cannot afford to become a test bed for new avionic design techniques except in a limited way.

Both of these factors have undoubtedly influenced the timing of the introduction of all-weather equipment but it now appears that an upsurge is occurring directed toward a wide-scale installation of reasonably conventional equipment. These equipments can initially, at least, operate to weather minima below the present 200' and half a mile, with a design aim to operate into the Category III regime.

Current Development Status of Systems for Operation Below Category II Limits

There are several 'all-weather' projects under active development today such as the Lockheed C141A program for Category IIIA, the Boeing/Sperry and Boeing/Bendix programs for the 727 and 707 series, the British Trident and VC10 programs. Several of these programs have produced systems which are capable of automatic touchdown under Category II weather conditions. Two notable firsts in commercial passenger service are the Trident automatic flare at London airport on June 1965 and the Pan-Am B-727 fully automatic landing at John F. Kennedy Airport on 27 February 1967. Both of these two systems are being further developed for eventual Category IIIA operation and represent a commendable step toward eventual operation under nearly zero visibility.

Despite this current, apparently encouraging progress, however, it is still not clear, especially in the United States what the eventual Category IIIA/B design concept will be especially in regard to the operational procedures and instrumentation requirements in the event of failures or unusual environmental conditions.

Objectives for All-Weather Landing Systems for Commercial Transport

In order to attempt a judgment on the real extent and value of our progress to date in the absence of a fully resolved technical requirement, the following objectives are suggested and the present state of development will be compared against these criteria. It is realized that these are very general but it is believed by the authors that they must all be satisfied to an acceptable degree if the final objective is to be achieved in an economic and orderly way.

- The system must meet the required safety criteria
- The system must possess adequate operational capability under realistic combinations of weather conditions
- The system design and configuration must be such that the flight crew are willing to use it under all prescribed operating conditions
- The performance and safety standards must be consistently achievable under routine airline maintenance conditions

Let us now examine our present capability and methods of achieving these objectives.

Safety

"The system must meet the required safety criteria." This statement can mean, in detail, different things to different people. The Air Registration Board in England has long declared the safety requirement as meeting a catastrophic landing accident rate of no greater than one in ten million landings - the famous 1 in 10^7. The ARB has written papers describing how it arrived at this criteria, but controversy still exists as to its validity. In general, the figure represents an aim at an improvement over the present accident rate of an order of magnitude and few people could argue that this must be the design goal.

The FAA has a general approach to specifying safety levels in terms of providing protection against failures which are not 'extremely remote'. Monitoring or redundancy is mandatory where the FAA feels it is necessary to maintain the safety levels to the highest practical limits. Advisory circulars are issued where relevant to provide groundrules for meeting the flight standards for the particular system involved. Whatever the specific requirement may be, the manufacturer must also set his own safety and performance standards and design accordingly.

What elements of the system are involved in the overall safety considerations? The answer in general terms is:

- The design and operation of the ground equipment
- The aircraft and equipment design and operation
- The operation of the aircraft and equipment by the crew

The design and operation of the ground equipment is the responsibility of the FAA and much work has been conducted concerning the redundancy levels of ILS ground transmitters, their siting, monitoring and susceptibility to environmental changes, especially disturbances due to traffic movement both in the air and on the ground. Despite intensive development and testing both in the United States and in Europe, a "cloud still hangs" over the quality of the signal as received by the aircraft and it appears that the vagaries of the ILS are still not completely understood. Development and testing still continues. To be fair, however, ILS installations consistently provide accurate guidance to thousands of airplanes down to present breakoff heights. The doubt is whether the present system designs can consistently provide the quality required when used as a vital guidance element for low visibility operations.

Most generally accepted appraisals of safety levels of the airborne elements of an 'all-weather' landing system consider only the airborne equipment in the following terms:

- The autopilot
- The sensors
- The main surface and trim servos or feeds to the power controls
- The cockpit selectors and switches
- The monitors and warning devices

Monitoring and/or redundancy is the accepted means of reducing the total system failure rate to a very low level. A failure fault and mode analysis would normally be carried out to increase confidence that the failure rates required will be met and to provide any necessary proof to certificating authorities. Failure analysis can be very time consuming if carried out diligently and it has been stated that the Trident all-weather program involved well over 10 man years of fault analysis on the subsystems involved which is still continuing.

In addition to the failure analysis of the airborne elements mentioned above, other associated equipments must also be considered such as the ILS and radio altimeter antenna, the power supplies to the equipment and of course, the primary controls of the airplane. With new airplane designs we would carry out a rigorous analysis of the total airborne system in order to insure a high safety level and also to optimize total system design in order to insure that there is not some inconsistent element which provides a 'weak link' in the safety chain.

The safety analysis by itself, of course, is not sufficient - the safety level must be preserved at all times especially in routine use of the equipment. Possible ambiguities in mode selection and switching or misunderstandings of warnings is a very important part of the design for very low failure rates. A comprehensive analysis of the design must be carried out in terms of pilot usage especially during the high stress period just before landing. Some of the techniques employed in a

visual flight simulator facility are described in a recent AIAA paper No. 67-399: "Developments Associated With Advanced Commercial Aircraft Crew Requirements."

Finally there is the insurance of the safety levels in service use. Although monitor and redundancy design can be verified to a high degree of confidence by preflight and inflight testing, the validity of the detailed fault analysis can only be preserved if all modifications to any part of the total airborne system are tracked and their effect on the original analysis reexamined. This aspect of safety preservation has had little consideration to date since the "see to land" concept of Category II operation allows for pilot intervention in case of malfunctions. Hence the level of failure, while required to be low, does not necessarily carry with it an attendant risk to the flight operations. When operating in visibilities where successful crew intervention is less predictable, then the inherent safety level of equipment must be maintained at a very high standard. In addition, of course, the validity of the statistical operational approach to confirm safety predictions by larger and larger samples of failure free operations can be completely invalidated by some change to the system which is not reconsidered in the context of the original failure analysis design studies.

Our plans to insure the safety criteria, apart from designing with the required level of redundancy and monitoring, will be to carry out the following programs for any Category III design.

a. Design the total vehicle from the outset for Category III operational safety levels.

b. Conduct a comprehensive check on design of all relevant vendor equipment and its relationship with the airplane design by formal means which will provide check and cross-check of all aspects of the all-weather equipment.

c. Carry out a formal program of fault simulation in the visual flight simulator to insure correct cockpit and systems design.

d. Conduct complete systems failure mode testing on the vehicle systems simulator (the iron bird) to verify systems design and the failure modes effects studies.

e. Verify the work done in c. and d. during flight test in a quantitive manner.

It can be seen that the program to design and check for failure mode effects which will not exceed the very low system failure rates is extensive and expensive. It must also commence early in the design phase of an airplane to have the best chance of success. Even then, as will be described later, the practical problems of crew confidence in the use of the system probably remains as a major element in the safety considerations.

Operational Capability

"The system must possess adequate operational capability under realistic combinations of weather conditions."

To obtain a high degree of touchdown performance accuracy, the automatic pilot must possess a control authority as great as the human pilot requires in all adverse conditions in which it must operate. This increased control authority requirement for operating to these lower altitudes is in direct opposition to the requirement of safety since a hard-over failure of such a system would leave the pilot little opportunity to recover - even in VFR conditions. Hence redundancy is required to permit the higher control authorities needed while preserving the safety of the operation in the event of malfunction.

What is adequate capability in terms of touchdown dispersions and what are the problems in achieving it? The evolutionary nature of the present system concept may possibly have led us into a trap with respect to attempting to produce adequate performance for "blind landing" from a system concept originally conceived to provide general guidance to around 200 ft. altitude so that the pilot could take over control and land visually. At 200 ft., moderate corrections to the flight path of the aircraft can be made. As the takeover altitude becomes lower, corrections become more difficult and, at about 100 ft., only very minor corrections can be tolerated; especially with large aircraft. For Category III conditions where fragmentary visual references begin to become available at 50 ft. or less, the aircraft position relative to the runway must be correct to a degree of probability such that the risk of landing outside the runway bounds becomes extremely remote. Let us examine the present control and guidance concepts that are presently universal for commercial automatic approach and landing.

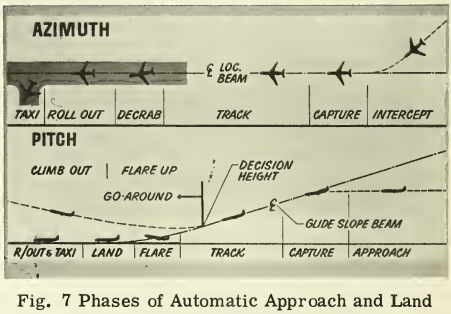


Fig. 7 Phases of Automatic Approach and Land

Figure 7 shows the phases of the landing procedure in the now familiar form and it is well known that the ILS beam is the sole guidance source down to approximately 50-70 ft. in pitch when flare commences. Flare guidance is provided by a radio altimeter signal which is computed on a rate of

descent demand prorated to altitude so that an exponential control to the runway surface is commanded. Pitch control is by means of elevators and trim mechanism. In azimuth, the ILS localizer beam is employed sometimes together with heading and computed to provide aileron steering down to the runway surface. Decrab can be manual in VFR, instrument guided or automatic in IFR conditions. Roll-out would normally be directed by signals derived from the localizer beam.

Now let us examine some of the problems facing the controls design engineer to produce the best possible accuracy from this system. First, what are the environmental influences which can affect system accuracy?

Errors Contributing to Landing Dispersions

Errors that contribute to the total landing dispersions fall into three categories:

1. Weather
2. Airborne equipment
3. Ground equipment

We are precluding pilot error in terms of an operational misuse of the system.

1. Weather - variations in wind direction, strength and constancy will obviously influence the landing accuracy in terms of lateral and longitudinal dispersions, vertical velocity, heading error and sideways velocity. These errors can be calculated on a probability basis on the assumption of maximum values of wind and assumed patterns of shear. Random gusts, however, or an unusual shear pattern may occur to cause a total error greater than expected. More work needs to be done in this area to increase confidence in the effect of wind variation on landing errors.

2. Airborne Equipment - the airborne contributions can be categorized as static (or datum), dynamic, and aircraft trim.

Static (or datum) errors can exist in any of the active airborne elements of the system such as ILS receivers and autopilot computers, and can exist for two reasons. The first will be a distribution of error around a nominal value due to factory test equipment and allowances which have to be made on a normal production line. The second will be variations of tolerances due to environmental changes, mainly temperature. Naturally a combination of errors must be statistically combined to assess the over-all effect and it is often difficult to accurately assess individual distributions especially those of manufacture and test.

Dynamic errors can exist as a result of tolerance variations in parameters, causing a change in over-all response of the automatic system. The error can either result in too tight a response, with possible consequent loss in damping, or too slack a response allowing larger over-all performance errors.

Aircraft trim variations in C. G. weight and approach speed can also produce position error in landing mainly as longitudinal dispersion or even touchdown velocity.

3. Ground Equipment - again errors in automatic landing can occur as a result of static (datum) errors or dynamic errors in the ground based ILS equipment. The static (datum) errors occur mainly as a result of error in aligning the beam and also because the ground based antenna monitors must allow datum variations to occur with a threshold if they are not to consistently cause failure alarms. Beam bends due to local terrain may also exist in certain sites, although these may be reduced as transmitter design improves. Errors can also be caused as a result of spurious noise on the received signal although filters can reduce this effect.

The effect of overflying aircraft can produce transient beam deflections and noise and this too may affect the touchdown accuracy of an automatic landing system.

In general, wind is the biggest contributor to touchdown errors with static, dynamic and trim all about equal. The effect of the former will probably never be completely predictable but improved knowledge of wind flow characteristics near the ground will help the designer to minimize its effects. Equipment errors, both ground and air, will decrease in influence as improved equipment designs and production techniques are developed.

Figure 8 shows the major control problems in the azimuth axis of control.

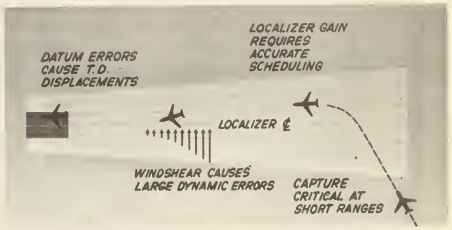


Fig. 8 Factors Affecting Azimuth Accuracy

Capture of the beam can only be commenced automatically by sensing the edge of the beam and initiating the "turn-in." This "edge" is only guaranteed within 2° either side of centerline although increasing air traffic is requiring captures at ranges as low as 4 to 6 miles. At these ranges

the distance in which the airplane must turn to capture the beam asymptotically is around 1000' and allowing for equipment tolerances and crosswinds, bracketing often occurs before the aircraft settles down. The existence of this activity so near the runway can influence the success of touchdown since the system does not have sufficient time to settle down on track. Better understanding of the system requirements for all-weather operation will help and the recent introduction of vectoring terminal area techniques may provide the guidance and logic to enable the airplane to turn in to the beam earlier at the closer ranges.

When established on the beam, the occurrence of wind shear, gusts and ILS noise all influence the system performance by varying degrees depending on the magnitude of these effects.

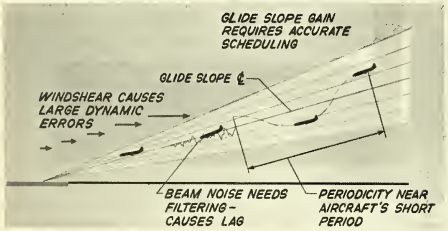


Fig. 9 Factors Affecting Pitch Accuracy

Figure 9 shows the pitch axis and some of its problems. Similar effects are experienced from winds and beam noise but here the situation is more critical because of the relative time response of the aircraft in pitch to the natural system period around the beam. Since the beam, as with the localizer, provides an angular error signal, its control effectiveness increases with decreasing range and a scheduling of this signal is required for stability reasons to maintain optimum control. This scheduling is arranged to decrease the system gain to nearly zero just before flare and the system is virtually open-loop at low altitudes, the very place where wind shears and gusts are most likely to occur. Also, because of the need to commence flare at a relatively high altitude to enable rate of descent reduction by elevators, the airplane is free from spatial guidance relative to the runway for around 12 seconds. Consequently, the longitudinal dispersions are higher than could be achieved with a "closed loop" down to the ground. The critical nature of control in pitch calls for a serious reappraisal of the present guidance and control techniques.

Figure 10 shows the typical touchdown performance requirements for an automatic landing system and it should be noted that these "requirements" represent the best that has been generally achieved by the particular system concept in use today.

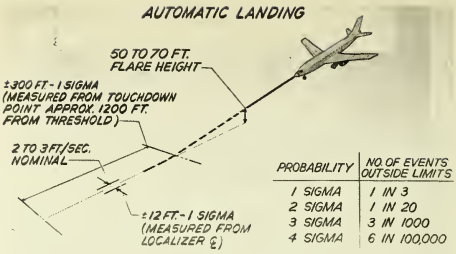


Fig. 10 Generally Accepted Touchdown Performance Standards

Figure 11 shows the longitudinal and lateral dispersion limits on a probability basis assuming a gaussian distribution of all the errors and disturbances involved. This assumption is probably reasonable because of the large number of influences many of which are random. It can be seen that on this basis the system may fall short of the original one in 10^7 criteria if left entirely alone on all landings since the wheels of a typical large airplane would be off the main runway on a number of occasions in ten million. It should be noted that the lateral dispersions shown do not take account of any offset in the ILS localizer beam centerline.

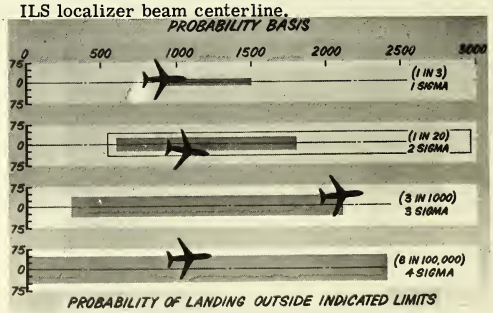


Fig. 11 Automatic Landing Touchdown Dispersion Limits

The question which arises is whether the touchdown performance illustrated is adequate operationally under all conditions of use including VFR conditions. The airborne equipments necessary to achieve this standard of performance are already considerably more sophisticated than those which the airlines are presently operating. It is doubtful whether this performance can be improved significantly by the present control techniques without a considerable increase in the sophistication of the system. Also, as airplanes become larger this will tend to offset any improvements affected by refinements of present system design concepts. One method of reducing the touchdown dispersion, of course,

would be to restrict the weather conditions under which the systems would be allowed to operate. Against this approach is the fact that insufficient knowledge exists of the extent and frequency of our low altitude weather conditions such as gusts, and wind shear and so the actual dispersions may occasionally be wider than the limits illustrated in Fig. 11. A decision is required by the potential users of automatic landing systems whether the present standards are adequate for regular use or is a significant improvement required.

Crew Confidence

"The system design and configuration must be such that the flight crew are willing to use it under all prescribed operating conditions."

It has been stated frequently that the automatic control system can produce a more consistent performance than the human pilot. In general terms, this statement should be true if we equate its performance against the total overall dispersions of a large sample of pilots in a wide variety of operational conditions at many different airports. However, one point which is overlooked in this generalization is that the average airline pilot is adaptable in that he takes account of expected and existing environmental conditions, his particular airplane configuration and perhaps his own stress level and optimizes his landing performance for each individual case. The particular runway in use obviously influences the landing he will perform especially in the case of short field lengths, wet runways, and early turnoff taxiways. Therefore, one fallacy in the comparison between the systems is that while the average pilot scatter may be greater than the average scatter of the automatic system the individual pilot performance represents an intelligent scatter predominantly aimed at performing an optimum maneuver to suit the circumstances and conditions that prevail at the time. Unfortunately the automatic system does not possess this intelligence and the scatter that it produces is an almost completely random one and in some cases it may bias the nominal touchdown point in the opposite direction from the one which the human pilot would choose. The automatic system would not, for instance, take account of a heavyweight airplane operating on wet runways and hence, tradeoff the risk of a touchdown point nearer to the threshold in order to balance the overall safety of the landing because of a possible problem in stopping distance. These factors represent some of the more subtle operational aspects of the use of the system in VFR conditions which the pilot will weigh against the advantage of overall better consistency in touchdown and the ability of the automatic touchdown system to land under non-visual conditions.

Now let us consider the case of the use of the systems under visibility conditions which are marginal or inadequate for the pilot to manually land the aircraft by visual means - Category IIIA and beyond. In order to be able to attempt to formulate this aspect from an overall systems engineering viewpoint, the following systems diagram has been prepared.

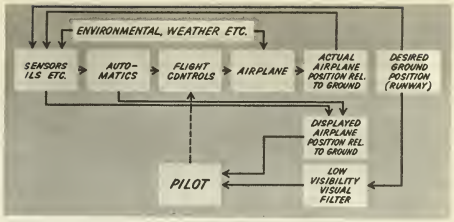


Fig. 12 Block Diagram of All-Weather Landing Functions

We can see that the guidance input of the system so far as achieving an actual ground position is concerned, is provided exclusively by the error signals from the ILS beam. This signal in combination with other sensing devices on the aircraft is computed in the autopilot which directs the controls of the aircraft to produce flight path deviations in response to the combined error signal created. The aircraft's actual spatial postion is thus produced and this needs to be monitored continuously by the pilot during his progress to the runway surface. His basic flight instrumentation provides him with deviation indications from the ILS beams together with basic flight attitudes, speed and rate of descent. At some point in the approach and landing - 100' in Category II conditions and probably 30 or 40' at Category IIIA conditions - he would obtain his first visual sight of the runway lights. Will this form of monitoring be sufficient for his confidence to remain at a level adequate to complete the landing in every case? And, if the first sight of the visual world conflicts with the mental picture of what he is expecting to see, will a correct operational judgment be made on a sufficiently high number of occasions especially under the high conditions of stress that may exist at this time? These are some of the questions which are being asked concerning the role of the pilot during an automatic landing in very low visibilities. Let us examine the present day type of instruments and displays on which the pilot is expected to monitor the approach accuracy.

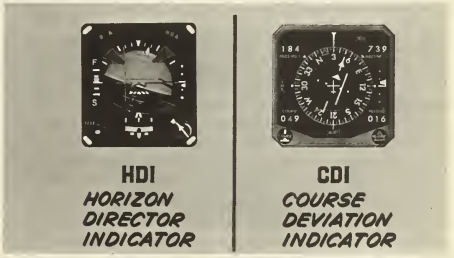


Fig. 13 Typical Situation Displays

It can be seen that the design of these instruments has evolved on the basis that the pilot can monitor all the important parameters within a very small

area of the instrument panel. This direction of design evolution has been influenced considerably by the operational procedures which call for a decision at 200' to continue head-upwith adequate sight of the runway. It is questionable whether the pilot would continue to monitor head-down to altitudes any lower than 200'. The present day interest in head-up displays has been promoted by the realization that a means is required which will enable the pilot to continue to monitor on instruments which will also permit him to watch for visual contact with the runway at the lower decision altitudes of lower weather minima. The next question that arises is whether, on first sight of the runway, the total information is complete enough to enable a firm and correct judgment to be made. Figure 14 shows the variableness of slant range visibility at low ceiling levels which have been shown to exist from experimental data originally gathered in Europe and which probably also exists at many major airfields in this country.

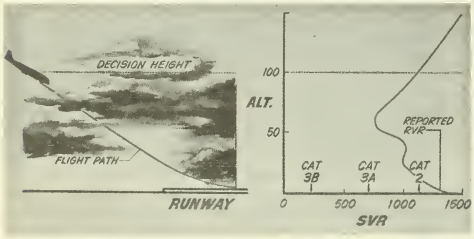


Fig. 14 Effect of Fine Grain Weather Variations on Slant Range Visibility at Low Reported RVR's

It seems inevitable that the pilot's decision on the reasonableness of proceeding below an altitude from which he can neither execute a go-around nor correct an inadequate positioning for safe landing will be based on outside visual cues if these are available. A clearcut transition from instrument guidance to adequate visual guidance even at very low altitudes could be tenable although the very short decision times must tend to degrade the quality of the decision.

The head-up display can assist by placing the relevant pilot's panel information in a position such that it can be continuously monitored up to the time when visual guidance becomes available thus improving the decision making process by eliminating the need for rapid re-accommodation of focus at a critical time. We are still left however with the problem of incorrect judgments due to misinterpretation of visual cues and to the very short decision times under conditions of high stress. A considerable improvement in the quality of the decision could be obtained if the decision time could be extended and the possibility of ambiguities of visual information decreased or eliminated. A high integrity monitoring means, which could blend into the real world as it comes into view, could achieve this objective. The decision to land or go-around could be made in a more relaxed manner. If the total system integrity could also be improved a system concept which is operationally acceptable to the average line-pilot would be attained. Figure 15 shows this concept in a systems block diagram format.

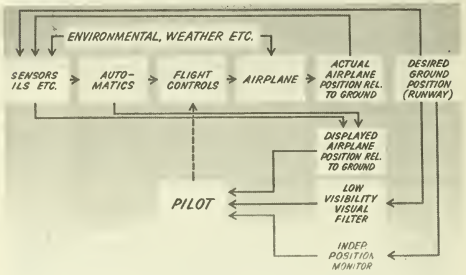


Fig. 15 Block Diagram of All-Weather Landing Functions

Independency appears to be highly desirable if for nor other reason than to decrease the systems reliance on the present sole guidance source - the ILS beams - the behavior of which still does not appear to be fully understood. There have been several devices developed such as Microvision which could fulfill this role of a "visual" monitor of runway position. Most of these systems have been designed for a long range capability of ten or more miles. The range requirement for the monitor considered in this present context need only be such that it becomes dependably operational from about 500 feet down - a range of about two or three miles. Thus, if the system uses a radar principle it would not require active transponders on the runway - a requirement which could impede progress because of the lengthy international negotiations and installation times that would ensue. Perhaps a more important consideration is that a system which does not require any significant ground installation would be immediately usable at all standard runways of the appropriate weather category anywhere in the world. It should be stressed that the independent monitor is not proposed as a means of landing but solely as a means of instilling pilot confidence which can improve the quality of the pilot's decision making process and insure the general safety level of landing in low visibility conditions. An illustration of the proposed head-up display presentation of the runway monitor and the appropriate flight progress information is deliberately excluded from this paper because this aspect needs careful development in conjunction with the pilots, human factors engineers and the aircraft manufacturer if it is to properly perform its task.

Basically the display could present in the approach mode combinations of the following parameters:

- Runway image superimposed on the actual runway

- Speed
- Radio altitude
- Roll and pitch attitude
- Rate of descent
- Autopilot mode
- Appropriate failure warning symbols

In our deliberations of the need to appraise head-up displays for approach and land monitoring we should not lose sight of the fact that the functional role of several of the present pilot's displays may now change. This fact accompanied by the recent developments in advanced display techniques invites a reappraisal of pilot's instrumentation with a possible reduction in his workload by this means.

Maintainability

"The performance and safety standards must be consistently achievable under routine airline maintenance conditions."

This aspect of all-weather operations is probably the least discussed of all. In one sense, it is probably one of the most important influences on the success of lowering present weather minima. We have already discussed the considerable increase in complexity that an all-weather system represents over present day approach systems and the varying degrees of success of these simpler systems. Maintenance skill levels are unlikely to increase in order to accommodate these new systems simply because of the lack of trained personnel in a rapidly expanding industry!

An operational need for a high systems availability will exist especially because of the competitiveness aspects of operating into airports where low visibilities exist for periods of time. In addition, the current unscheduled removal rate of avionic and associated equipment is high with at least 50% of the units removed reported by the maintenance shops as "no fault found." Quite obviously, an improvement in this situation is required if all-weather landing is to earn its keep. Undoubtedly the airlines will adapt their maintenance techniques and facilities within reason to accommodate these new systems as they have many times in the past, but the main effort in insuring a high usability and a reasonable maintenance cost must come from the manufacturers facing up to the challenge of producing systems designs compatible with the capabilities of the airline maintenance system. It is self-evident that all-weather and other sophisticated avionic systems being introduced concurrently will add to the maintenance load of the airlines. The deleterious effects of this additional load must be minimized by sound intelligent design of both equipment and testing techniques if the systems concepts are to be acceptable to the airlines for day to day in-line service use. It would be ironic if the introduction of a system originally intended to improve regularity of service actually degrades it because of a poor maintainability record which would increase operating costs and decrease system availability and safety. If this system is rejected soon after introduction for any of these reasons, progress to lower weather minima operations will be seriously delayed.

Conclusions

Although some attempt has been made in this paper to cover all the aspects concerned with the introduction of all-weather landing into airline service use, the authors are very conscious of having omitted several factors from the total considerations that need to be made. It is hoped, however, that the areas covered in this paper represent the most important ones which need consideration if we, as an industry, are to make a success of another milestone in history of commercial transport aviation.

All-weather landing has perhaps been one of the most debated subjects over the past decade and probably more papers have been written on this than on any other subject in the general field of airplane subsystem design. An attempt has been made to honestly present today's situation and to highlight some of the major problem areas that need consideration. There is no intention to disparage all the progress that has been made and the sincere efforts in many quarters toward the earliest introduction of this increased capability for our air transport system. On the other hand, we believe that because of all the progress that has been made it is now necessary for all interested parties in all-weather landing to consolidate the many lessons we have learned into a worthwhile end result. Many of the papers written during various time periods over the past few years have emphasized that all-weather landing is with us now - yet this does not appear to have happened. Recently, articles have appeared which are critical of the general progress in this field. That there is additional work to be done has already been discussed in this paper. We believe, however, that this work may be summarized under the following headings and that a resolute effort to conduct these activities will enable us to move progressively but rapidly into regular operation under low visibility weather conditions. The areas that need attention are as follows:

a. The evolutionary nature of the development of automatic flight control systems has lead to touchdown performance results which we believe may not be adequate for day to day use of the automatic touchdown capability. Any improvement in control techniques which can improve this situation would be a worthwhile contribution.

b. The role of the human pilot in all-weather operations has not been adequately

resolved. He has complete responsibility for the flight and he will only delegate control of the aircraft to an automatic system if he has full confidence in its performance. The head-up display appears to offer considerable promise in this respect and the development of a simple independent monitor for use with this display may well remove many of the pilot's doubts of the operational capability of the total airplane/system/pilot combinations. We believe that the pilot's role must be resolved between the pilots and the airplane/systems designers before further significant progress to lower weather minima operations can be made.

c. Two design aspects of the airborne equipment must receive close attention.

The first is that with the use of redundancy to maintain safety levels, the safety analysis of the total equipment involved in the all-weather landing operation must be rigorously applied and monitored.

The second is that, in view of the questionable performance of automatic approach couplers up to the present time in airline service use, a special effort must be made to insure that the routine performance of all-weather landing systems is improved considerably if we are to expect the pilot to be prepared to use the system in flight regimes where safety can be hazarded by erratic performance.

d. The maintainability aspects of advanced avionic systems must be continuously developed to insure that the somewhat erratic record for today's approach couplers is not repeated.

As a manufacturer, we intend to adopt an overall systems approach to the incorporation of a genuine all-weather landing capability and will make the best possible use of all the scientific resources available to us to insure an integrated systems approach to the airplane design, the equipment design and the pilot's requirements. We believe that we are entering an era of operational dependency on complex systems which demands a more scientific systems approach if we are to most efficiently and rapidly optimize the design, operation and maintenance aspects concerned. Further delays in achieving all-weather operations will be caused not by an insurmountable technical inability but rather by a lack of communication in sufficient depth between all parties concerned. This lack of communication is one that can be corrected. There is no doubt in our minds that all-weather landing can be achieved safely and within a reasonable time period.

-- NOTES --

-- NOTES --

No. 67-573

AUTOMATIC LANDING SYSTEM 727 AIRPLANE

by

J. E. TEMPLEMAN
The Boeing Company
Renton, Washington
and
R. H. PARKER
Sperry Phoenix Company
Phoenix, Arizona

AIAA Paper
No. 67-573

AIAA Guidance, Control and Flight Dynamics Conference
HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

2.09, 7.11, 8.08

-- NOTES --

BOEING/SPERRY AUTOMATIC LANDING SYSTEM 727 AIRPLANE

J. E. Templeman
The Boeing Company
Renton, Washington

R. H. Parker
Sperry Phoenix Company
Phoenix, Arizona

ABSTRACT

A milestone in the history of American commercial aviation was reached on the night of 27 February 1967. A Pan American Boeing 727 jetliner made a fully automatic landing at JFK International Airport in New York. This was a scheduled flight with 98 fare-paying passengers on board. With visibility restricted by snow flurries, and despite crosswinds gusting to 12 knots, the touchdown was smoothly executed within 4 feet of runway centerline. This relatively unheralded feat was the culmination of years of effort by many related groups in the field of aviation.

This paper considers the changing role of the automatic flight control system in commercial aviation relative to function and safety. It describes the evolution of the 727 system with respect to its use in obtaining progressively lower operational weather minimums for the airlines. The overall landing system is then described, explaining details of the dual-pitch channels of the automatic pilot, and the unique methods used to implement a system that is both safe and dependable. Following a discussion of the 727 automatic landing certification program, the future role and form of the automatic flight control system are assessed.

INTRODUCTION

The role of the automatic flight control system (AFCS) in commercial aviation had not, until recently, changed very dramatically since the first automatic pilot was introduced into fleet operation nearly 35 years ago. It is true that AFCS capabilities have grown markedly with the addition of various modes and functions. This is indicated in Figure 1, which lists their functions by airplane axis. The early models of AFCS did nothing more than maintain a fixed airplane attitude for the pilot. The typical AFCS found in commercial transports today provides most of the functions listed in Figure 1. However, the role of the AFCS did not change; it was not a dispatch item, but served primarily to reduce pilot workload and add to the comfort of the passengers. In this capacity, the safety aspects of the AFCS were passive in nature. That is, it was necessary to demonstrate only that the design of the AFCS was such that it did not in any way compromise the safety of the airplane. The interface between the automatic and manual control systems was designed so that the automatic system could be readily and positively disconnected from the manual controls in the event of a malfunction of the automatic controls. The control authority of the automatic system was limited so that, no matter what the malfunction, it could not subject the airplane to excessive maneuvering loads or cause attitude excursions from which the pilot could not safely recover.

LATERAL/DIRECTIONAL CONTROL

- YAW DAMPING
- ROLL ATTITUDE HOLD
- HEADING HOLD
- TURN KNOB MANEUVERING
- PRESET HEADING CONTROL
- VOR RADIAL TRACKING
- ILS LOCALIZER TRACKING

LONGITUDINAL CONTROL

- PITCH ATTITUDE HOLD
- BAROMETRIC ALTITUDE HOLD
- PITCH KNOB MANEUVERING
- AIRSPEED HOLD
- ILS GLIDE SLOPE TRACKING
- AUTOTHROTTLE CONTROL

FIGURE 1. AFCS MODES/FUNCTIONS

With the development of the commercial jet transport, the AFCS was called upon to perform some new functions, and for the first time its role began to change. (See Figure 2.) The swept-back wing of the jet transport accentuated the lateral directional instability known as "dutch roll". The safety of an airplane is directly related to its handling qualities, and less than satisfactory handling qualities resulted from the dutch roll instability. It was highly desirable to augment the characteristics of the basic airplane with a means for damping the dutch roll. As a matter of fact, in some commercial jet transports, this damper is mandatory for safe operation in some portions of the altitude/airspeed flight envelope. In effect then, a portion of the AFCS has become a dispatch item.

- PROVIDE PILOT RELIEF AND PASSENGER COMFORT
- ENHANCE SAFETY BY IMPROVING HANDLING QUALITIES
- ALLOW OPERATION AT LOWER WEATHER MINIMUMS

FIGURE 2. EVOLUTION OF AFCS ROLE

Another phenomenon demonstrated by the high speed jet transport is that known as Mach tuck-under. This effect manifests itself as a divergent

67-573

mode whereby an incremental increase in speed (Mach number) causes a nose-down pitching moment on the airplane, which in turn causes a further increase in speed. Although the amount of Mach tuck-under, and the particular range of Mach number over which it occurs will vary from one model of airplane to another, it typically occurs in the region of Mach number 0.8 to 0.9. This phenomenon again represents a compromise of handling qualities and is of such a magnitude in some models as to be considered unsafe. The AFCS was again called upon to come to the rescue and to provide automatically a compensating pitching moment that would eliminate the divergent mode. This is typically accomplished by programming elevator deflection as a precise function of Mach number over the critical Mach range. Again, a subsystem of the AFCS was required to provide the margin of safety deemed necessary for high speed operation of commercial jet transports.

Thus, we find that the role of the AFCS relative to safety is changing from passive to active, and two of its subsystems have already become dispatch items.

It is not only the handling qualities of the jet transport that have acted as stimuli in changing the role of the AFCS. The rapidly increasing density of air traffic in major terminal areas has provided the incentive to find the means by which it will be possible to operate into and out of these terminal areas in spite of adverse weather conditions. The economic penalties and personal inconveniences resulting from delayed arrivals and diversions to alternate airports are becoming intolerable.

The task of defining acceptable means by which operational weather minimums may be safely reduced, has proven to be monumental. Notwithstanding the massive effort already put forth, there is still much to be done before "zero-zero" or all-weather operation can be realized. One thing has become apparent: the lowering of operational weather minimums must be accomplished in realistic steps. The steps have been defined by the FAA, with the help of the rest of the aviation industry, in terms of weather minimums categories. The equipment requirements and the associated performance and safety criteria for each of these categories have been only partly defined. Briefly, the weather minimums categories are as shown in Figure 3.

One of the major deterrents to defining the equipment and facility requirements for lower operational weather minimums is the lack of operational experience in those environments. Boeing and Sperry recognized this at the beginning of the 727 program and made certain key decisions which would help to resolve this problem. It was decided that in taking each of the steps toward lower minimums, we would impose design objectives that, in our best judgment, would provide the capability for operating to the next lower weather category. We felt that only by doing this would the pilots be able to gain the operational experience so vital to defining the associated performance and safety criteria for that next category. Let us trace this progression to date, with reference to Figure 4.

CATEGORY	MINIMUM DECISION ALTITUDE (FEET)	RUNWAY VISUAL RANGE (FEET)
I	200	2600
II	100	1200
IIIA	0	700
IIIB	0	150
IIIC	0	0

FIGURE 3. FAA WEATHER MINIMUMS CATEGORIES

CATEGORY	CERTIFICATION DATE	MINIMUM ALTITUDE (FEET) IFR	VFR
I	OCT 1963	200	78
II	OCT 1965	100	78
II WITH AUTOMATIC LANDING	JUN 1966	100	0

FIGURE 4. CERTIFICATION HISTORY OF 727 AFCS

The AFCS originally certified in the 727 was designed to meet the safety and performance requirements for CAT I, but with sufficient margin to provide stable and accurate path control on the ILS localizer and glide slope beams down to altitudes below 100 feet. Furthermore, its elevator authority was limited so that a malfunction would not result in significant altitude loss which would otherwise limit its use to some altitude above 100 feet. As a result, in October 1963, the 727 AFCS was certified for clear weather operation down to an altitude of 78 feet. By demonstrating performance and pilot proficiency, the airlines soon obtained FAA authorization to operate under CAT I weather minimums conditions and, having established visual contact at the 200 feet MDA (minimum decision altitude), to continue under automatic control down to 78 feet. The operational experience gained through this and other programs led to the formulation of criteria for CAT II operation. These new accuracy criteria required that certain improvements be made to the 727 AFCS, primarily in the glide slope control circuits of the automatic pilot. This improvement program led to a recertification of the 727 AFCS in October 1965 for CAT II operation, and these improvements were delivered to the airlines. By early 1966, through demonstrations of system performance and pilot proficiency, the airlines began obtaining FAA approval to operate in CAT II weather minimums conditions.

The step to CAT IIIA operation was not as simple. First of all, performance and accuracy criteria were not defined, to say nothing of safety requirements. However, one hurdle which was defined and had to be overcome was the so-called 50-foot rule. That is, the autopilot could not be used below an altitude corresponding to 50 feet plus the amount of altitude lost during manual recovery from a nose-down hard-over malfunction of the autopilot. Since the manual recovery must allow for a 1-second fault recognition time by the pilot, it was obvious that the autopilot had to be made "fail-passive". This, and other practical considerations, led to self-imposed design criteria which can be rather simply stated as follows:

> The pilot shall be immediately warned of a malfunction, and no malfunction shall disturb the flight path of the airplane to the extent that the pilot cannot take over manually and complete the landing safely or execute a go-around.

Furthermore, consistent with the philosophy of providing operational margin below the CAT IIIA objective, it was deemed necessary to provide automatic flight path control throughout the flare maneuver to touchdown. By certification of such a system, we would again be in a position to obtain airline operational experience which would aid in defining the criteria for a further reduction in operational weather minimums.

The following paragraphs describe the system that was implemented in the 727 to provide the automatic flare function and to realize fail-passive characteristics.

727 AUTOMATIC LANDING SYSTEM DESCRIPTION

The equipment which has been added to the 727 to enable certification of the airplane for automatic approach in CAT II weather minimums consists of improved ILS receivers, a radio altimeter and indicator for minimum decision altitude determination (MDA), an improved rain repellant system, and autopilot modifications to provide tight ILS beam tracking. Equipment necessary for the 727 automatic landing system beyond that required for a CAT II airplane is as follows:

a. Second radio altimeter

b. Second autopilot pitch channel

c. Autopilot flare coupler (dual)

d. Autopilot roll monitor channel

e. Autopilot approach progress display modifications

This information is given to provide a feel for the equipment configuration difference between a CAT II and an automatic landing airplane. Following is a description of the elements that make up the Boeing/Sperry-developed automatic landing system.

In the pitch axis of control, two pitch control channels are used to independently drive separate elevator hydraulic actuators. Figure 5 is a simplified block diagram of the dual pitch system. Complete duality is maintained in the pitch axis system from the input sensors to the elevator control servos, inclusive. The only inter-tie between the two pitch control channels is an equalization signal path. The equalization signal has limited authority, and it enables the two channels to track within system accuracy requirements in the presence of normal component tolerances. This equalization feature is discussed in more detail in a later section of the paper. A dual flare coupler provides the flight path computation between the glide slope control phase and touchdown. The flare coupler commands an exponential flare profile as a function of altitude rate and displacement, and is programmed to control the touchdown rate of descent to 2-1/2 feet per second. Radio altimeter information is used during both the glide slope and flare modes of autopilot operation. The altimeter signal is used in the glide slope mode to program the glide slope gain as a function of altitude above the terrain, with the unique feature that depressions in the terrain that appear as an increase in altitude are ignored. In the flare mode, the altimeter signal is the primary source for the flare coupler command computation. It is evident that the radio altimeter plays a key role in the automatic landing system, and development of reliable accurate units by the radio altimeter manufacturers was timely.

Tracking of the two pitch channels is continuously monitored by fail-safe amplifiers. If a malfunction causes the tracking error to exceed a predetermined amount, the pitch channels disconnect automatically.

The roll axis of control is based upon a dual concept, but the implementation is somewhat different from that of the pitch axis. In this case, the second channel is a roll monitor that duplicates the computations of the roll channel, but it does not drive a separate aileron control servo. Figure 6 is a simplified block diagram of the roll control and monitor system. Fundamentally, the roll monitor is a comparison-type monitor. It receives signal information from a separate set of sensors, and operates on these signals to compute aileron commands. It compares these commands with the corresponding command signals in the roll channel to verify their integrity. Having validated the command signals, it also computes what the response of the aileron servo should be, and compares this computed response

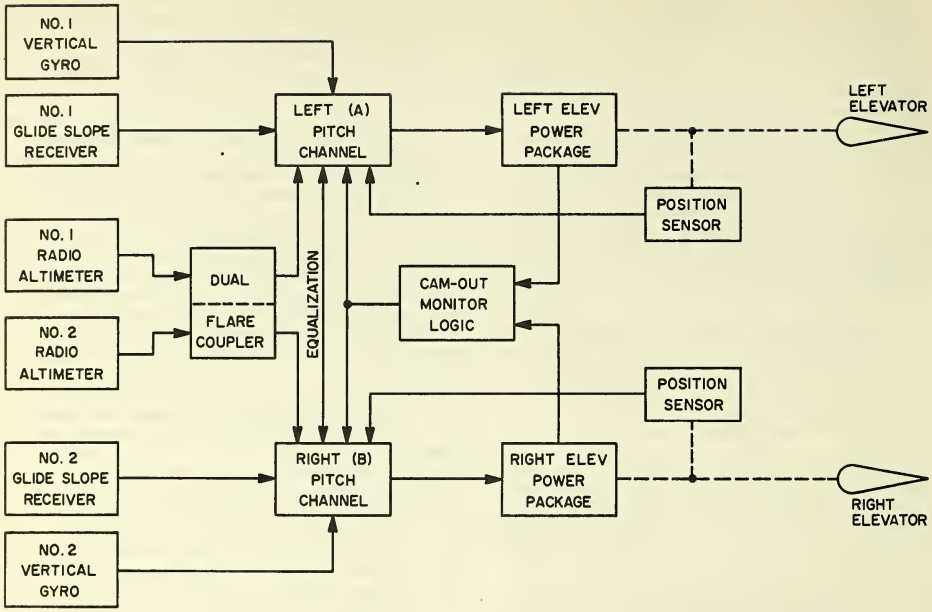


FIGURE 5. DUAL PITCH AXIS CONTROL SYSTEMS

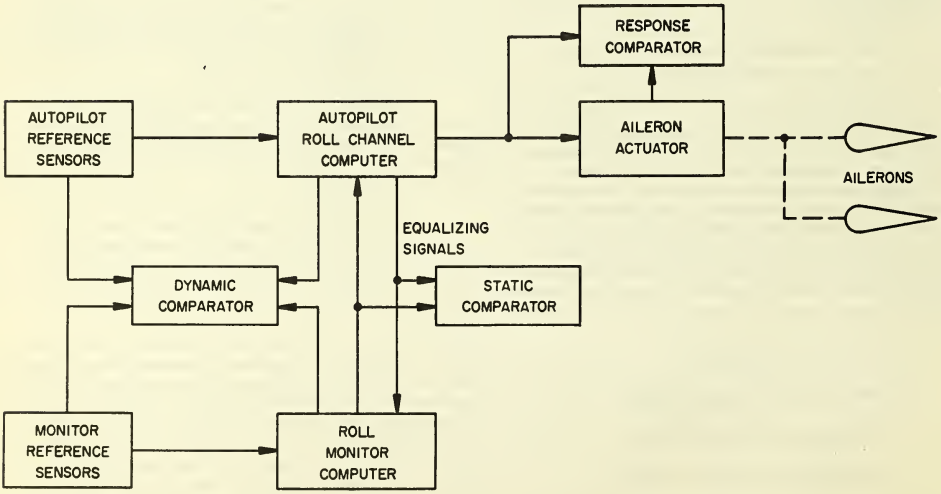


FIGURE 6. ROLL CONTROL AND MONITOR SYSTEM

with that of the actual servo. The various signal comparisons are effected by a fail-safe amplifier; if either static or dynamic errors exceed prescribed thresholds, automatic disengagement occurs, with pilot warning. This system also utilizes the equalization technique to minimize the effect of tolerance build-up in equipment. The fail-safe amplifier is of the same type used in the pitch axis and will be discussed in more detail in a later section of this paper.

The autothrottle system used on the 727 airplane was developed by Boeing and Sperry to specifically enhance the performance of the automatic approach and landing system. In reducing the pilot's workload, it also allows more time for his role as the master monitor of the automatic approach progress. The autothrottle, however, is not a required subsystem of the 727 automatic landing system. The function of the autothrottle system is to maintain a preselected airspeed by automatically controlling the engine throttle levers. Figure 7 is a simplified block diagram of the autothrottle system. Upon engagement, the autothrottle system operates the thrust levers to control the airplane indicated airspeed to coincide with the airspeed indicator speed command "bug" which is manually set by the pilot. The airspeed error signal is combined with a longitudinal accelerometer signal through a low-pass filter to minimize throttle activity in gusty air conditions. A pitch attitude signal is used to augment the airspeed error computation to reduce transient airspeed errors during normal airplane maneuvers. The autothrottle system complements the automatic flare maneuver by retarding the throttle levers at a fixed rate during the flare mode of autopilot operation.

The approach progress display is an array of annunciators used to inform the pilot of the autopilot mode of operation. Figure 8 illustrates the sequence of events during an automatic approach to touchdown and the associated indication displayed to the pilot. Six cues are given by amber and green indications from three light annunciators. An automatic approach to touchdown begins with the autopilot engaged and the airplane course set to intercept the localizer. The automatic coupling mode is selected and the following events take place:

a. The localizer annunciator will be amber, indicating that the roll axis control system is in an armed state for automatic acquisition of the localizer beam.

b. The glide slope annunciator will be amber, indicating that the pitch axis control system is armed for automatic acquisition of the glide slope beam.

c. Upon intercepting the localizer beam, the localizer annunciator turns green, and the roll control system initiates capture of the localizer beam.

d. Upon intercepting the glide slope beam, the glide slope annunciator turns green, and the pitch control system initiates capture of the glide slope beam. Up to this point in the approach, only one pitch channel has been engaged. Upon intercepting the glide slope beam, the second pitch channel is engaged automatically, and fail-safe monitoring begins.

e. At approximately 1500 feet of altitude, the autopilot initiates localizer and glide slope gain programming in response to radio altimeter signals. Simultaneously the roll monitor is activated, which in turn arms the autopilot flare mode. This is indicated to the pilot by the flare annunciator turning amber. The approach is now, and will remain, fully monitored to touchdown.

f. At approximately 50 feet of altitude, the pitch axis control system switches from the glide slope mode to the flare mode. Initiation of the flare mode is indicated by the flare annunciator turning green. Normal time between flare initiation and touchdown is approximately 8 seconds.

DUAL PITCH CHANNEL DESCRIPTION

In order to implement a fail-passive system, it is necessary to employ a means for reliably detecting both active and passive failures, and a certain amount of duality is required, particularly at the sensor level.* Once the sensor signals are validated, the response of the remaining components to the signals can be validated by comparison monitoring techniques. Alternatively, the system can be completely dualized so that a failure of one channel is counteracted by the opposing channel. As indicated by the previous description of the system, the autopilot of the 727 AFCS uses both techniques. The dual simultaneously operating pitch channels are considered unique and will be explored in more detail in the following paragraphs.

The 727 basic elevator control system contains dual electrohydraulic actuators. Although a comparison monitor was considered for the pitch channel of the autopilot, there were two significant advantages to implementing the dual channel system, namely, airline logistics would benefit by having two channels of identical interchangeable components, and dual channels afforded a potential easy growth to a fail-operational system.

*In some cases, equivalent reliability may be obtained by in-line monitoring of a single sensor, but some sensors have failure modes that cannot be practically covered by this technique.

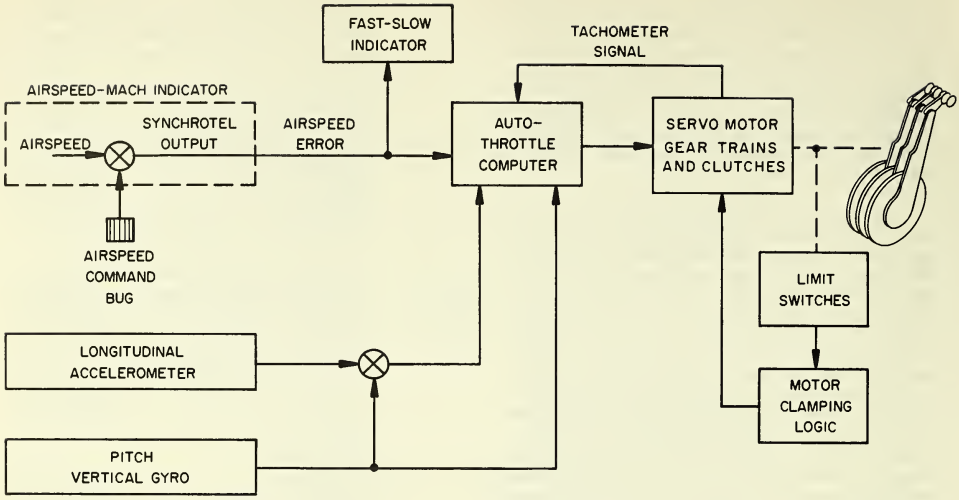


FIGURE 7. 727 AUTOTHROTTLE SYSTEM

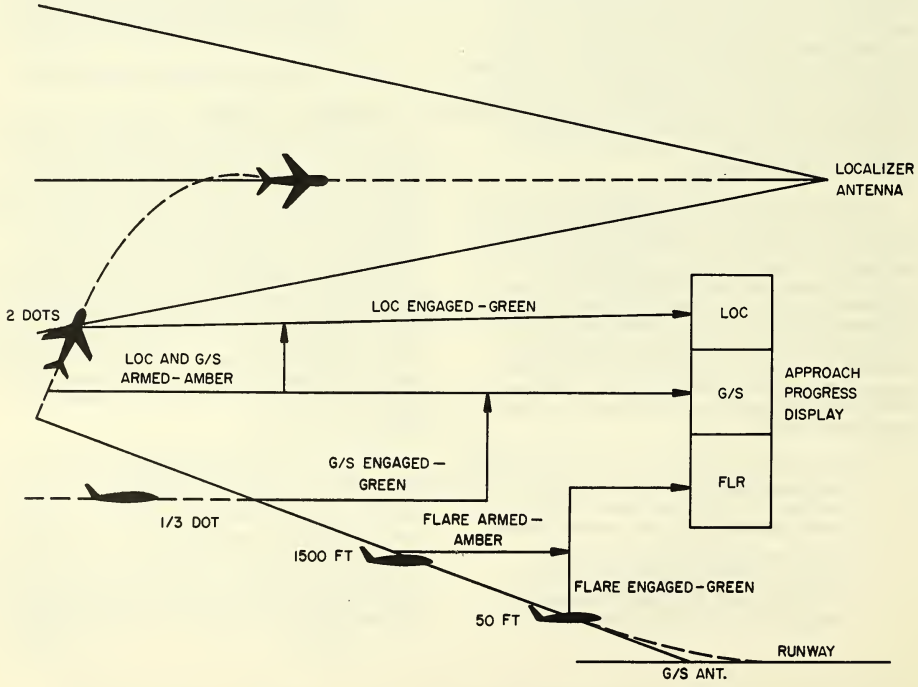


FIGURE 8. APPROACH PROGRESS DISPLAY

To fully appreciate the fail-safe protection of the 727 dual pitch channels, it is necessary to investigate the elevator actuators, control system linkages, and associated artificial feel system. This is most clearly done with reference to the mechanical schematic diagram of Figure 9. For the sake of simplicity, the actuators are shown as single-stage devices driven by a pilot's control valve (for manual operation) or by an electrohydraulic transfer valve (for automatic pilot operation). The feel system is shown as a simple mechanical spring. The 727 elevator actuators are actually two-stage devices, and the feel system varies the effective feel spring gradient as a function of flight conditions.

Referring to Figure 9, it is seen that the right and left elevators are not bussed together, but they are connected together through the actuator input linkages of the control system. Each elevator is powered from a separate actuator. In manual operation, when the pilot moves the control column, the control cables rotate the main quadrant. The feel spring provides an opposing force consisting of an initial preload plus an additional force proportional to quadrant rotation as it compresses the feel spring. The rotation of the quadrant is transmitted through the bell cranks and linkages, and displaces the pilot's control valves with respect to the actuator housings. As a result, hydraulic pressure

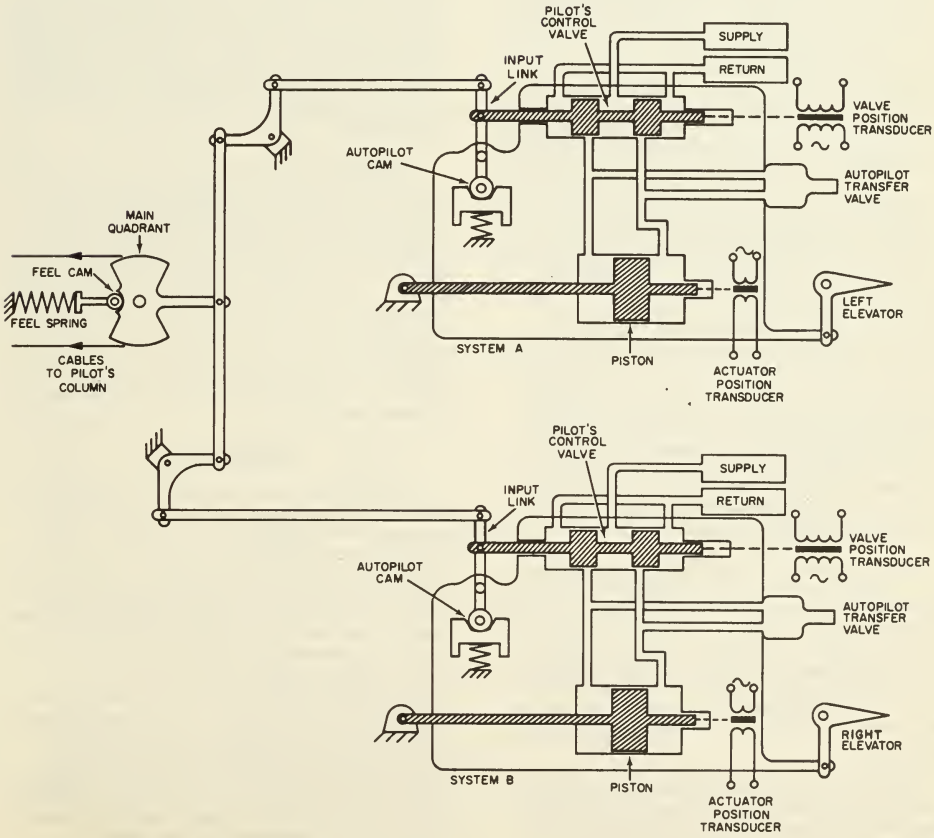


FIGURE 9. ELEVATOR CONTROL SYSTEM

67-573

is ported to the main pistons, causing motion of the actuator housings and corresponding deflection of the elevators. Motion of the actuator housings results in recentering of the pilot's control valves.

In single-channel autopilot operation, hydraulic supply pressure is connected to one autopilot transfer valve. Also, the input link that drives the pilot's control valve of the same actuator is caged by a spring-loaded cam (autopilot cam); this prevents movement of the pilot's control valve with respect to the actuator housing. When the autopilot amplifier supplies dc current to the transfer valve, the valve ports hydraulic pressure to the main piston, causing motion of the actuator housing and a corresponding elevator deflection. As the actuator moves, it forces the control system linkages to follow it, thus rotating the main quadrant and compressing the feel spring. Note that the pilot's column follows the movement of the actuator, and that the linkages of the opposite actuator cause the pilot's control valve of that actuator to move. It can be seen then, that the opposite actuator is "slaved" to the autopilot-driven actuator, and both elevators deflect an equal amount. The actuator position transducer provides a feedback signal to the autopilot amplifier that cancels the command signal at the amplifier input and rebalances the transfer valve when the elevator has reached the commanded position. If a large command signal is present at the amplifier input, the actuators will move and deflect the elevators up to the point where the reaction force of the feel spring is greater than the spring preload force at the autopilot cam. Then a small increase in actuator displacement will cause the input link to rotate (since the autopilot cam spring compresses), and the pilot's control valve will move. The mechanical advantage and relative valve authorities are such that a small motion of the pilot's control valve will completely cancel the hydraulic input from the autopilot transfer valve, and further actuator displacement is prohibited. Thus, it is seen that autopilot elevator authority is determined by the feel spring gradient relative to the autopilot cam spring preload.

Next, consider the dual-pitch channel operation. In this case, both actuator input links are caged by their respective autopilot cams, and each autopilot transfer valve is driven from a separate autopilot channel. As long as the two autopilot commands agree, both actuators (and elevators) will respond together. Also, since the reaction force of the feel spring is distributed equally between the autopilot cams, the autopilot authority is doubled.

Assume now that the autopilot commands do. not agree; a typical case to consider would be one where both systems are initially balanced, with the elevators faired. If a command is transmitted to system A transfer valve only, then the A actuator will start to move, and in so doing it will attempt to rotate the main quadrant as well as the input link to system B actuator. However, to do so it would have to overcome the preload force of the feel spring plus the preload force at the B autopilot cam. Since the sum of these preload forces is greater than the preload force on the A autopilot cam, that cam spring will yield, allowing the input link of A actuator to rotate and move its pilot control valve. This control valve input cancels the transfer valve input for a very small elevator deflection. The airplane will start to respond to the small elevator deflection, but as it does so, the sensors of the B autopilot supply an opposing command to the B transfer valve. As the B actuator tries to move and rotate the main quadrant, the preload forces of the main quadrant and the A autopilot cam must be overcome. Again, the sum of these forces is greater than the preload force on the B autopilot cam, and its cam spring yields. This allows the input link to the B actuator to rotate and move its pilot control valve. Now the input from the B transfer valve is cancelled; the two elevators are slightly deflected in opposing directions, and the airplane continues on its path undisturbed. The dual system is seen, therefore, to be inherently "fail-passive."

In the following discussion the phrase "cam-out" will be used with reference to an actuator, when its autopilot cam spring preload has been over-ridden and its transfer valve flow is cancelled by flow from the pilot's control valve. Also, it will be seen in Figure 5 that linear transducers are connected to the pilot's control valves. The polarity and magnitude of their signal outputs are indicative of pilot valve travels, and supply intelligence when cam-out occurs, as well as the direction of commanded surface which caused the cam-out. These transducers will be referred to as "cam-out sensors."

EQUALIZATION

In operation of dual autopilots, it is to be expected that tolerances will be such as to result in unequal commands to the two transfer valves. This does not affect performance, as it can be shown that this system will respond to the smaller of the two commands, with one actuator controlling. The other actuator will be cammed-out and slaved to the controlling actuator. However, the cam-out transducers provide intelligence to a failure-warning system, and nuisance failure warnings cannot be tolerated. Therefore, means were devised for equalizing the commands from the two autopilot channels. Figure 10 shows a simplified block diagram of this equalization. There is a certain amount of compliance in the control system linkages, so that relative elevator positions can disagree by approximately 1 degree before the force unbalance of the system is enough to initiate cam-out of an actuator. An equalization reference signal is obtained by subtracting the two actuator position transducer

signals. This error signal is fed back to the two autopilot channels in a sense to increase the command to the elevator of least deflection, and decrease the command to the elevator of greatest deflection. The equalization signal is fed back to the output amplifier of each channel for short-term correction, and to an integrator in the signal path for long-term correction. The "authority" of the equalization signal is very important since it tends to propagate a failure from one autopilot channel to the other channel. With the implementation chosen for the 727, the equalization signal authority is inherently limited at the sensor level by the mechanical compliance of the system. (The elevators can never disagree in excess of the compliance range without causing actuator cam-out.) Gain authority is controlled by using redundant passive circuit elements whose failure modes cannot result in a gain increase. Hot-wire fault protection is inherent in that a fault at any point in the signal chain will cause dual opposing actuator cam-outs which has been shown previously to be a passive failure. The authority of the equalizer signal is restricted then to only that required to compensate for tolerance build-up in the two pitch channels, and thus keep the elevators tracking within the compliance range below the cam-out threshold. For example, in the 727 a hardover failure at one autopilot output amplifier will result in dual opposing cam-outs of the actuators for a pitch attitude transient of less than 1 degree, and automatic disconnect will occur in less than a second, warning the pilot and leaving the flight path essentially undisturbed.

It is worth pointing out here that a most desirable goal has been achieved through the previously described implementation. The components of the system can be maintained within practical tolerances and will provide dependable performance which is both safe and free from nuisance warnings. It is not enough for a system to be reliable in terms of the probability of detail part failures, and safe in terms of fault detection. If it is necessary to maintain the components to unreasonably close tolerances in order to preclude nuisance warnings and disconnects, then such a system will not be dependable and will not contribute to lowering the operational weather minimums.

MONITORING OF THE DUAL PITCH CHANNELS

It has been shown that the dual pitch channels, operating into the isolated hydraulic actuators of the 727 control system, have inherently fail-passive characteristics when considering an active failure in either channel. However, in order to implement a completely reliable system, it is necessary to consider not only active but passive failures in both the hydraulic actuators and in the autopilot control channels. An exhaustive failure analysis has indicated that all modes of failure are manifested by the relative states of only four signals in the dual channel system. The signals of significance are the two cam-out transducer signals and the two signal outputs from the autopilot output amplifiers. Certain relative states of these signals are indicative of

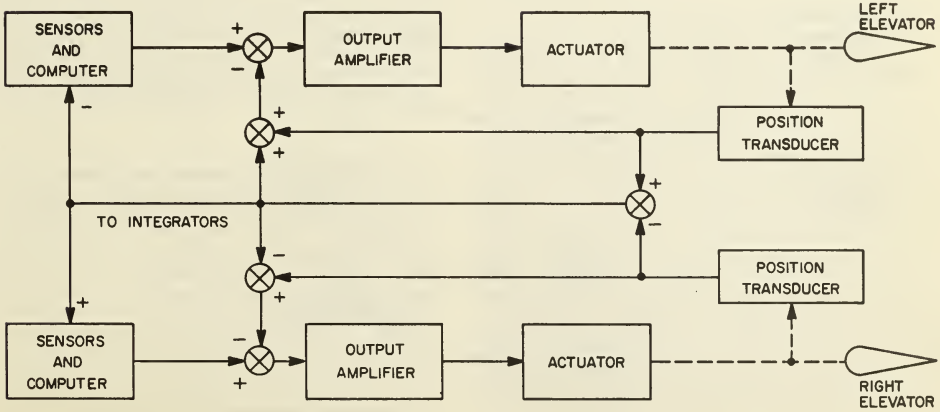


FIGURE 10. DUAL PITCH CHANNEL EQUALIZATION

67-573

malfunctions whereas other relative states are indicative of normal performance. Fail-safe monitoring, with automatic disengagement and pilot warning, required the transposition of the states of these four signals into corresponding states of redundant relays, in a fail-safe manner. The mechanism for achieving this transposition has been referred to as a fail-safe amplifier. (See Figure 11.) Under normal operation, the output of this amplifier energizes two pairs of relay coils. In order that these relay coils remain energized, it is necessary that the amplifier properly transmit an 800-cycle signal, applied at its input. In this way, the fail-safe amplifier is self-checking. A signal that is to be monitored is connected to the input of the fail-safe amplifier. If the level of the monitored signal exceeds a prescribed threshold, it will cancel alternate pulses of the 800-cycle reference signal and result in de-activating the associated pair of output relays. The pair of output relays which is de-activated corresponds to the phase of the monitored signal. The integrity of the monitored signal path is established by the use of tracer signal techniques. The foregoing is not intended to be a failure analysis of the monitoring system, but rather to indicate the basic philosophy used in monitoring of the dual pitch channels. By transposing signal states into corresponding states of redundant relays, and then by interconnecting the contacts of these redundant relays in the proper fashion, the monitor provides automatic disengagement in the event of a malfunction in the system. At the same time, it ignores those conditions resulting from normal operation.

The redundant relay contacts of the fail-safe monitor are interconnected to perform the following functions:

- Disengage the system if a single cam-out is indicated unless the cam-out agrees with the command from the autopilot output amplifier.

  This protects against a failure in the actuator which would cause it to revert to the slaved mode, or a failure of the transfer valve (active or passive).

- Disengage the system if dual opposing cam-outs are indicated.

  This protects against any active or passive failure in the sensors and electronics or an active failure of the actuator.

It is significant to note that the monitor ignores a single cam-out when it agrees with the sense

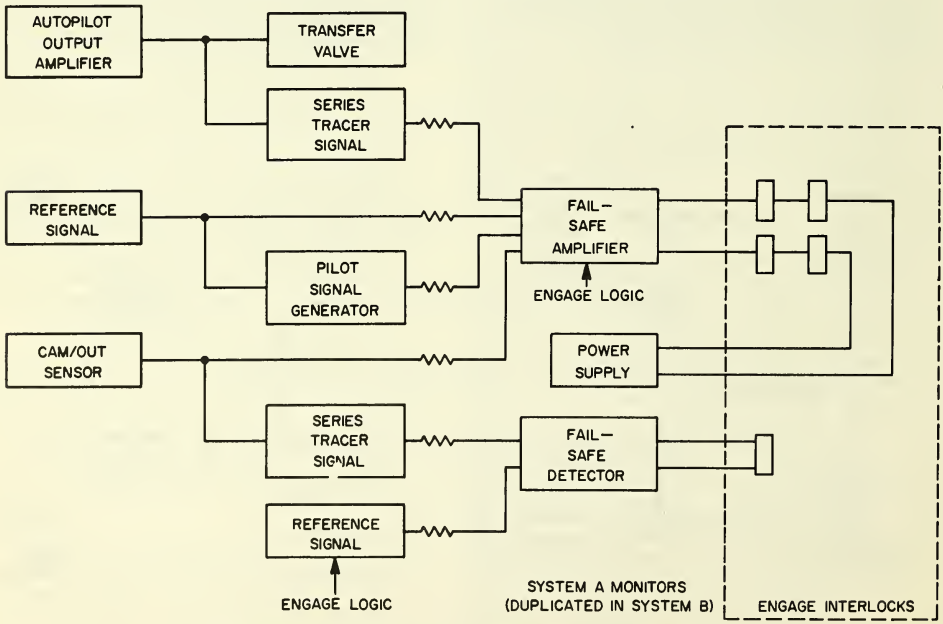


FIGURE 11. DUAL PITCH CHANNEL MONITORING

of the autopilot command, as well as a dual cam-out if the cam-outs are in the same direction. The first of these might occur for a short period because of tolerance accumulations and does not warrant a disengagement; the second might occur for a short time depending upon the magnitude of trim required when flaps and gear are extended at glide slope capture. It merely indicates that the autopilot has temporarily reached its elevator authority limit and is not cause for disengagement.

It is seen then that the monitor has met the objective of detecting true failures and eliminating nuisance disengagements, thereby providing both the safety and dependability required for lower weather minimums.

CERTIFICATION PROGRAM - LABORATORY PHASE

Prior to flight testing the fail-passive automatic landing system, laboratory tests were carried out with the autopilot hardware in a simulated airplane environment. The test setup consisted of coupling the preproduction autopilot hardware to the 727 Flight Controls Test Rig and an analog computer simulation of the 727 airframe equations of motion. The objectives of the laboratory tests were to evaluate the effect of system tolerances on both performance (including safety) and dependability (susceptibility to nuisance failure alarms) and to verify the monitor responsiveness to simulated malfunctions.

System failure modes investigated in previous autopilot programs involved only passive or instantaneously saturated malfunctions of signal paths. Since the automatic landing system is operating in one of the most critical phases of flight, the failure detection monitor was required to detect time dependent failures (ramp type malfunctions); e.g., a slowly changing vertical gyro reference caused by a failure in its erection circuit. The failure mode investigation included taking into account those tolerance distributions that would be most significant in reducing the monitor sensitivity for detecting each failure condition.

The expense involved in demonstrating all failure modes of the system in flight tests was prohibitive. Therefore, the laboratory test results were combined with a detailed description of the monitoring techniques to form a "failure analysis" document of the automatic landing system. This document was submitted to the Federal Aviation Agency (FAA) along with a proposed flight test demonstration specification document in request for an FAA-type inspection authorization. The flight test specification called for demonstrating the failure modes which the laboratory study showed to be most critical, as well as demonstrating the system automatic touchdown performance.

A Type Inspection Authorization (TIA) for the Model 727 Automatic Approach and Landing System was issued by the FAA, dated 28 December 1965. Basically, the TIA approved testing of an automatic approach and landing system configuration to meet performance requirements of CAT II weather conditions and automatically land the airplane under VFR conditions below the CAT II weather minimums. The automatic landing performance was to be evaluated using best known criteria at that time without benefit of established approved criteria by the FAA. At Boeing, the TIA approval set into motion the final phase of a new system development: flight demonstration to the FAA.

FAA FLIGHT TEST EVALUATION

In order to demonstrate the automatic landing system to the FAA, the autopilot had to be in a production configuration, and the test airplane had to reflect production standards for all equipment that interfaced with the autopilot. The availability schedule of Boeing's 727 flight test airplane and preparation of the final paperwork necessary to describe the installation of the system on the airplane delayed formal system presentation to the FAA until 3 March 1966. The TIA required that specific ground checks be made and witnessed by FAA personnel, to substantiate that the automatic landing system installed on the test airplane conformed to the drawings and diagrams presented to the FAA in request for a TIA. Bench and airplane ground tests also substantiated that the monitor threshold levels were at their least sensitive values with respect to malfunction detection, and that the control system power packages reflected the worst case tolerance situation established by the system failure analysis for the most critical system failures.

In-flight demonstration of the system to the FAA began on 15 March 1966. All simulated malfunction testing and the majority of performance testing was conducted at the Oakland International Airport in California. The ILS facility there was published to be of CAT II quality. Performance testing was also conducted at Portland International Airport in Oregon, at Seattle-Tacoma International Airport in Washington, and at Boeing Field in Seattle to provide cross-comparison data of the system operation at different ILS ground facilities. Ninety percent of the FAA demonstration flight tests, which involved 72 approaches, were completed after 3 days of flying. The FAA required that satisfactory system performance be demonstrated under moderate turbulent, windshear, and crosswind air conditions. These air conditions could not be found prevailing at an acceptable local airfield, but after approximately 5 weeks of watching weather reports in increasingly larger radial distances from Seattle, the desired

conditions were found at Stapleton International Airport, Denver, Colorado. The flight tests were completed there on 29 April 1966.

Thirty-four approaches were conducted at Oakland with simulated malfunctions. All malfunctions were inserted at altitudes below 100 feet (CAT II MDA). At no time was the FAA pilot made aware whether a given approach was to demonstrate performance or to demonstrate the system responsiveness to a failure. The types of malfunctions demonstrated were as follows:

a. Autothrottle advance and retard hardovers during flare.

b. Elevator nose-down and nose-up hardovers at approximately 80, 50, and 30 feet of altitude.

c. Elevator nose-down and nose-up ramp fault commands at approximately 80, 50, and 30 feet of altitude.

d. Failure of the system to provide a flare command.

e. Aileron hardover at flare.

f. Roll vertical gyro ramp fault at approximately 80 and 50 feet of altitude.

g. Both yaw dampers inoperative and one operative with a hardover at flare.

The ramp-type malfunctions were inserted both above and below the monitor threshold. At no time did the FAA pilot respond to an inserted malfunction by visual cues before the autopilot monitor disconnected the system. Most approaches were accomplished with the pilot remaining under the hood well below the CAT II MDA, and in some cases, below 50 feet. Recognition of a system malfunction by a disconnect of the autopilot was followed by a manually controlled go-around or landing at the discretion of the pilot. Failure of the flare mode was demonstrated at the request of the FAA to assure that the airplane remained in a nonhazardous attitude following fade-out of the glide slope control mode.

Sixteen approaches were conducted at Oakland to demonstrate the automatic approach and landing system performance to touchdown. All met the CAT II performance criteria to 100 feet, and were continued to touchdown. Figure 12 illustrates the main gear longitudinal and lateral touchdown dispersion relative, respectively, to a line on the runway designating the glide slope antenna location and the runway centerline. Included in Figure 12 are the touchdown points for approaches where simulated malfunctions went undetected or were detected by the autopilot monitor just prior to touchdown. As previously mentioned, no specific FAA criteria existed with which to compare the performance results of the automatic landing system. However, the flight test results were compared with the results of laboratory studies, and all results agreed within 15 percent of predicted error values.

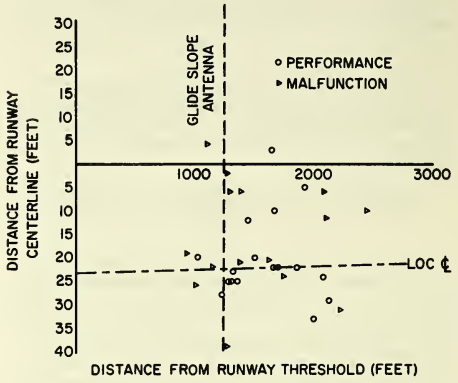


FIGURE 12. TOUCHDOWN DISPERSIONS RUNWAY 29 - OAKLAND

The approaches conducted at fields other than Oakland, to demonstrate system performance at several ILS facilities, exhibited comparable touchdown dispersion to those shown in Figure 12. The Denver approaches satisfied the wind conditions of moderate turbulence, as classified by the FAA pilot, and were conducted with crosswind components between 13 and 26 knots as calculated from the airport tower reported winds. All approaches whether for performance or malfunction detection, were considered by the FAA pilot to provide the controllability desired of an automatic approach and landing system, with an adequate margin of safety when the autopilot automatically disconnected upon recognition of a malfunction.

The FAA demonstration flight test data was reviewed, compiled, and documented in a flight test summary report and then submitted to the FAA for final approval of the system tested. Type certification for the Boeing-Sperry 727 automatic approach and landing system was received from the Federal Aviation Agency in June, 1966.

CONCLUSIONS

As far as Boeing and Sperry are concerned, the 727 automatic approach and landing system met the self-imposed design goals established at the outset of the program. Industry and the FAA have yet to detail an acceptable performance criteria for CAT IIIA automatic landing systems.

However, an Agency Advisory Circular draft was issued this spring specifying the performance criteria for automatic landing under CAT II weather minimums. In all probability this circular will, with minor modifications, be the basis for performance criteria for CAT IIIA automatic landing systems.

Let us now compare the results of the 727 performance testing at Oakland with the FAA published criteria for automatic landing under CAT II weather minimums. The FAA circular states that the aircraft centerline and main gear touchdown dispersion should be shown on at least a one sigma probability to lie:

a. Laterally, within ±20 feet of the ILS localizer beam center

b. Longitudinally, within -300 and +900 feet of the intersection of the glide slope beam center and the runway

Figure 13 shows the performance touchdown data, presented in Figure 12, replotted with respect to the Oakland ILS localizer beam center. For the sample shown, the one sigma distribution falls well inside the FAA boundary. In fact, the FAA boundary can well represent a two sigma distribution. However, note that this data reflects only a small sample of touchdowns, and an accumulation of in-service data must be gathered under normal operational environment in order to substantiate any conclusions drawn here.

The 727 automatic landing system developed by Boeing and Sperry represents the AFCS state of the art as it exists today. The system has the capability both in performance and safety to extend the weather minimums to CAT IIIA. As new techniques are developed in the area of "fail-operational" approach systems, whether the philosophy be complete automatic control or specific roles for both pilot and autopilot, the 727 system has the potential for growth to an all-weather blind landing system.

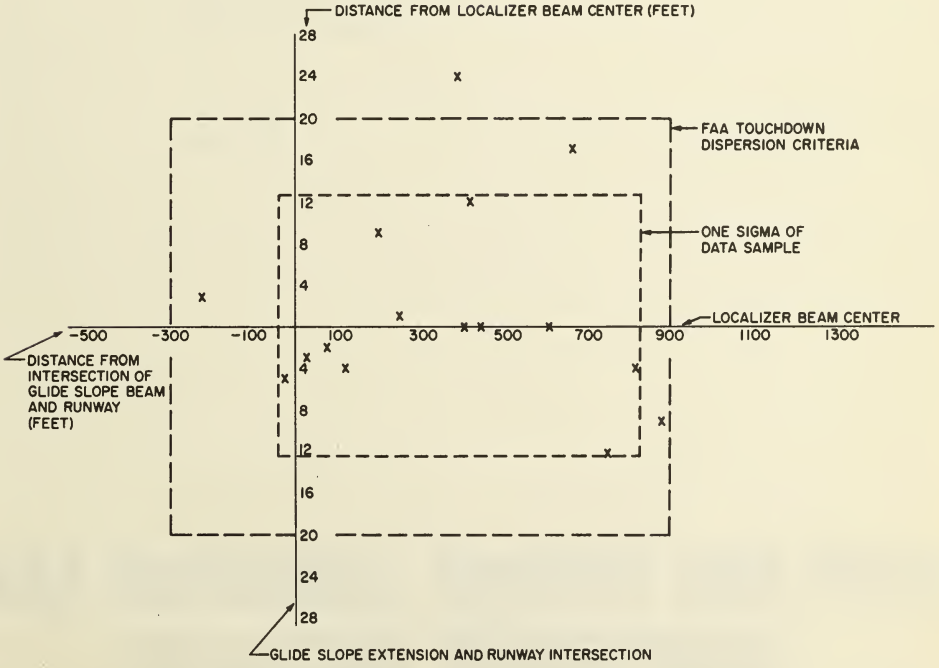


FIGURE 13. 727 AUTOMATIC LANDING SYSTEM CERTIFICATION TOUCHDOWN DISPERSIONS COMPARED TO FAA CRITERIA

67-573

The role of the AFCS, as shown by the 727 system presented, adds a greater degree of safety to the critical approach and landing phase of flight. It relieves the pilot of the strenuous task of controlling the airplane, allowing considerable more pilot attention to be focused on monitoring the approach from an overall safety standpoint.

A sharper definition of AFCS and pilot roles in controlling the airplane can now be visualized. The AFCS, with its self-monitoring logic, controls approaches in an accurate, smooth, safe, and repetitious manner. The pilot monitors the general performance of the approach relative to his ILS information display, with specific instructions on what action is required when the AFCS indicates a system malfunction. The ability of the AFCS to detect equipment failures and provide "fail-passive" operation eliminates the dangerous situation of a pilot being required to execute a recovery maneuver following an active equipment malfunction. The pilot's role becomes one of continuing, where the AFCS left off, path control of the airplane to complete the approach or initiate a go-around.

To extend the commercial airliner's lower weather minimums operation to the CAT IIIB and C levels, new self-imposed design objectives are being established. In short, some type fail-operational system will be necessary. Along with this is the realization that ground guidance is needed to assist the pilot in directional control during the landing roll-out. These areas are being worked on by industry for the coming generation of commercial aircraft.

No. 67-585

AUTOMATIC ACQUISITION OF CANOPUS

by

G. B. DALY and L. T. SEAMAN
General Electric Company
King of Prussia, Pennsylvania

AIAA Paper
No. 67-585

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

AUTOMATIC ACQUISITION OF CANOPUS

G. B. Daly and L. T. Seaman
Missile and Space Division
General Electric Company
King of Prussia, Pa.

ABSTRACT

A means for uniquely acquiring the star Canopus for roll control of interplanetary spacecraft is described. The identification function is completely autonomous after launch, and would thus reduce the ground command and control requirements for future missions, such as Voyager. Discrimination is made solely on the basis of observed source "brightness", and is implemented through a pair of acquisition gates set closely above and below Canopus' level. Ratios of gating levels to Canopus' brightness are maintained through periodic in-flight calibration against a stable light source contained within the sensor. Performance requirements are defined through an investigation of potentially acquirable objects, including the brightest stars, planets and planetary satellites. Sources of gating level variation which affect the accuracy of in-flight calibration are examined, and their combined effect determined in an error analysis. Typical sensor response to bright regions of the Milky Way is determined to infer the probability of triggering false acquisition on these areas. Alternate means of providing unique automatic Canopus identification, including spectral discrimination and pattern recognition techniques, are discussed briefly. A demonstration model sensor has been assembled and operated.

INTRODUCTION

The Sun and the star Canopus form a convenient reference system for attitude control of translunar and interplanetary spacecraft. While acquisition of the Sun is readily accomplished, unique identification and acquisition of Canopus is considerably more difficult even though it is the second brightest star in the sky. Systems currently in use have proven adequate, but only with auxiliary monitoring by the ground station to verify that the acquired star is Canopus, and to override the system if it is not. This paper describes a sensor which will uniquely identify and acquire Canopus without assistance from earth-based equipment or personnel.

Such a sensor is of particular value to interplanetary spacecraft such as Voyager, now planned to be launched to Mars in 1973. After each of its midcourse corrections, after the Martian orbit insertion, and after any orbit trim maneuvers, the Voyager must reacquire its attitude references. Automatic acquisition of Canopus can avoid potentially difficult periods of ground station operation, and increase the overall system reliability, during each of these periods. Once the spacecraft "locks" on the Sun, it rotates about the Sun line (the roll axis) searching for Canopus. The Canopus sensor will view many stars of almost the same brightness as Canopus during the spacecraft rotation. The particular stars which pass through the field of view will depend upon the mission phase since typical Voyager trajectories, from the first midcourse correction to achievement of a satisfactory Martian orbit, carry the spacecraft roughly 180 degrees around the Sun. If the sensor can not distinguish between Canopus and these other stars, the spacecraft may lock on an incorrect star. An operational procedure must then be followed to recognize this false lock and, by earth command, to override the control system to release the star and move on to the next. Eventually, Canopus will be found. Since there will be two Voyager spacecraft launched for each Martian opportunity, this might create a severe operational problem if both spacecraft should need simultaneous earth-based assistance in acquiring Canopus. A Canopus sensor that can automatically identify and acquire Canopus is a better solution to this problem.

METHODS OF STAR IDENTIFICATION

To uniquely identify and acquire Canopus, one or more characteristics of that star must be positively recognized. Such observable characteristics include:

a. Position with respect to other stars
b. Spectral Irradiance
c. Intensity

A listing of these characteristics is given in Figure 1 for the six brightest stars as seen by a detector having an S-4 spectral response. This abbreviated list will aid in illustrating some of the problems involved in successfully identifying Canopus.

a. Position With Respect to Other Stars

Canopus can be recognized by virtue of its relative position in the sky in several ways, depending upon the number of star sensors being used.

An excellent example of a multiple sensor acquisition system is the set of six star trackers used on the Orbiting Astronomical Observatory (OAO). Prior to roll search about the sun line, each tracker axis is rotated to a unique pointing direction, relative to the spacecraft. Simultaneous detection of the corresponding set of pre-selected guide stars during the search then guarantees acquisition of the required reference system.

A far simpler version can be devised for interplanetary spacecraft such as Voyager by using a set of just two star sensors to simultaneously view Canopus and Vega. These two stars are very bright and are nearly opposite one another in the sky, as shown in Figure 2. Thus

the probability that acquiring these stars establishes the required roll reference is very high. The design of a Vega star sensor is relatively straight forward since it should be possible to modify the basic design of the flight-proven Mariner IV Canopus sensor; however, the cone angle range must be doubled to accommodate Vega's 28 degree offset from the North ecliptic pole. Since this method does require an additional sensor, it can decrease the total reliability of the mission, and so a simpler method was sought.

A second technique based on a combination of position and intensity data requires only a single sensor. In this method, the sensor output during roll search is recorded and compared with star position and brightness data stored in a computer. A correlation between actual and stored data is then made, and the spacecraft is rotated through the correct angle to bring Canopus into the field of view.

A ground-based version of this method was used by JPL during the historic flight of Mariner IV to Mars in 1964.[1] A predicted sensor output-vs-roll angle plot was generated in a ground-based computer and compared with the sensor output signal telemetered from the spacecraft. The comparison is shown in Figure 3.

A variation of this approach which might be useful for Voyager requires that the star atlas data be stored in computer memory on-board the spacecraft. Since this storage is greatly restricted in capacity, only data on the brighter stars would be available. Thus only a partially complete "predicted" plot could be generated which complicates comparison with the actual sensor output. Another problem is that of predicting which stars located near the edge of the sensor's field of view will be recorded, since the actual search path can vary from the nominal one due to the attitude control limit cycle in pitch and yaw. These difficulties were considered sufficient cause to reject the on-board computer approach, although ground-based "map matching" would almost certainly be used to verify Canopus acquisition.

b. Spectral Irradiance

Differences in spectral irradiance are indicated in Figure 1 by the star classification and apparent color. Depending on this observable characteristic alone for a unique star identification is not attractive because a sophisticated sensor is required and, due to the similar spectra of several of the stars, very poor discrimination is provided. However, use of this characteristic in combination with others can emphasize star differences and make identification easier. For example, the measured star brightness or intensity is a function of the spectral response of the detector used in the star sensor. Thus by using transmission filters to modify the nearly identical spectral response of available phototubes (S-4, S-5, S-11, S-20), it is possible to increase the brightness difference between two stars.

For example, simplified relative spectra of Canopus and Vega are shown in Figure 4.[2] The difference in energy between these two plots is greater in the visual region (50% response at 5100 and 6100 Å) than in the S-11 region (50% response at 3500 and 5500 Å). This fact was illustrated in Figure 1 by the corresponding differences in visual and S-4* magnitudes of $0^m.73$ and $0^m.58$, respectively. Thus if filtering is used to change the sensor spectral response from S-4/S-11 to visual, the intensity difference can be increased by $0^m.15$. This is a small gain, however, and is achieved only through a great decrease in overall sensitivity, which may make the sensor S/N ratio unacceptable. Thus a larger receiving aperture and phototube may be required, resulting in major redesign of the Mariner IV type sensor and the image dissector phototube. These modifications are considered unjustified. Other techniques of measuring spectral radiance were not considered in depth during this study due to the complexity of design, potential unreliability, and the data reduction associated with multi-band photometric sensors.

c. Intensity

Finally, differences in star intensities may be used to uniquely identify and acquire Canopus. From Figure 1, it is seen that, to an S-4 or S-11 sensor, Canopus falls between Sirius and Rigel in brightness, and that Vega is nearly as bright as Rigel. Since Rigel and Vega are closer in brightness to Canopus, they are the more difficult to reject. Furthermore, as mentioned previously, Vega is nearly opposite Canopus in the sky and thus will often be "seen" during roll searches.

To assure reliable unique acquisition of Canopus by this method, celestial sources other than stars must also be rejected. The only objects in the Solar System which may enter the sensor field of view and which may appear near Canopus' brightness during a mission to Mars are the two Martian satellites and Saturn.[3] The former can have the same brightness as Canopus only for a very brief period near encounter. Preliminary calculations have shown that Saturn can reach Vega's brightness at periods of maximum ring exposure. More rigorous investigation is required to determine its actual maximum brightness. Since these celestial sources can interfere with automatic acquisition during only short and completely predictable periods of the mission, it is felt that they are acceptable deviations from a completely effective automatic Canopus acquisition system. All other planets and satellites are either brighter than Sirius, fainter than Vega, or cannot appear in the sensor field of view.

It is conceivable that bright regions in the Milky Way could appear as bright as Canopus in the sensor's relatively large instantaneous field of view (10 x 0.85 degrees). The Mariner IV flight sensor indicated a response of only 10% of Canopus' level as it crossed the galactic plane in the constellation Puppis.[1] Figure 5 shows this point at galactic longitude 222 degrees and reveals an integrated sky brightness

*Nearly identical to S-11.

of 500 stars of $10^{m}.0$ per square degree. [4] Since the maximum is seen to be 700 stars, the sensor's response to the brightest region in the Milky Way should only be about 15% of Canopus' level and will therefore be easily rejected.

d. Selected Star Identification Method

From the considerations discussed above, it is concluded that, if a sensor can be devised with a sufficiently narrow "brightness acceptance range", discrimination solely on the basis of intensity is the simplest and therefore probably the most reliable means of uniquely acquiring Canopus. The remainder of this paper discusses a means of implementing such a discrimination system, and the status of our present effort to experimentally verify the practicality of the approach.

REQUIREMENTS FOR INTENSITY DISCRIMINATION

Figure 6 is a list of potentially acquirable stars, showing their S-4 brightnesses relative to Canopus. [5] To maximize the probability that Canopus will always be recognized, and Rigel and Vega will always be rejected, the lower brightness threshold or gate should be set at 0.80 times Canopus. An upper gate set at 1.50 times Canopus will reject Sirius and all other bright sources. Clearly the lower gate is the more critical, and its threshold must be maintained with an error less than 0.20 times Canopus or 25% of the gate level.

In a star sensor, these gates are typically implemented by comparing a small fixed percentage of the detector high voltage (which varies inversely with star brightness) with a reference voltage in a threshold detector. Input voltages below the upper reference voltage (indicating a sufficiently bright source) generate a logic signal that is used to enable the sensor. Voltages below the lower reference voltage (indicating an excessively bright source) generate a logic signal that is used to inhibit the sensor. The accuracy of the gate setting is therefore affected by the accuracy of the high voltage supply circuits (accuracy of the automatic gain control loop) and of the threshold detectors. The 25% stability requirement can not be met "open loop" with assurance of long-term, dependable operation due to component degradation and gain changes with time, temperature, and radiation effects. For these reasons, as shown in Figure 6, the Mariner IV Canopus sensor was operated with the recognition gates set rather far from Canopus (0.12 to 0.30 times Canopus for the lower gate and 8 times Canopus for the upper). [1] The Mariner system, therefore, could (and did) acquire several stars other than Canopus.

INTENSITY DISCRIMINATION WITH IN-FLIGHT CALIBRATION

The method chosen to maintain the recognition gate levels within the desired limits is that of periodic in-flight calibration. The proposed method is based on use of a stable light source which is imaged on the detector photocathode outside the normal field of view. The light level of this source is set to produce the same photocurrent as Canopus. Periodically, the sensor scan is offset briefly to view the calibration source image on the photocathode. The normal and offset scan patterns are shown in Figure 7. The resulting calibration signal is used to establish and store in digital form a source voltage for the upper and lower reference levels to eliminate gate variations between calibrations. The reference levels are thus dependent upon all loop gains at the time of calibration or "up-dating". In normal operation, the actual input signal level resulting from the star field being observed is compared to these temporarily stored reference voltages to decide if Canopus is present in the field of view. Those factors which affect how nearly equal to Canopus' intensity the gates can be set are discussed in the section on calibration error analysis.

The in-flight calibration equipment has been designed as an addition to the basic Mariner-IV Canopus sensor, although the concept can be applied to other designs. Figure 8 depicts the automatic sensor block diagram, with new or modified components distinguished from the original Mariner IV design.

As the calibration cycle begins, the scan pattern in the image-dissector is deflected to the image of the calibration source (see Figure 7). The source produces the same photocurrent as Canopus and, through operation of the automatic gain control (AGC) loop (including the 1 & 2 KH_z bandpass filter, peak detector, AGC reference voltage generator, comparator, and high voltage control amplifier and supply circuits) establishes a voltage near 1 KV across the electron multiplier in the image tube.

The output of the high voltage control amplifier is a convenient indicator of the intensity of the observed source and in Mariner IV was telemetered as a "brightness signal". [1] In the calibration system, this voltage is stored in an A to D to A converter. The converter chosen is of the "ramp" type, which essentially counts high rate clock pulses until the stored level reaches the value of the input voltage. The counter generates a quantized ramp voltage through a resistance ladder connected to analog switches on each bit in the counter. The total expected variation in this "intensity voltage" due to gain changes in the AGC loop is only a small percentage of its nominal value. Thus only the additional voltage increment above a base level (somewhat below the expected minimum) will actually be placed in digital storage. This eliminates useless re-counting of non-changing voltage increments, thus reducing the total number of bits and the counting time required.

The voltage stored is equivalent to Canopus intensity, and reflects the system gains at the time of calibration. Sufficiently frequent calibrations will assure that gain changes with time, temperature, and radiation will be adequately compensated.

To generate reference voltages for the upper and lower gate comparators, a feedback-stabilized amplifier is used in conjunction with a precision voltage divider. The amplifier raises the stored voltage to the level corresponding

to 0.8 times Canopus (voltage and brightness are inversely related) and the voltage divider provides the lower voltage which represents 1.5 times Canopus. By this technique, the AGC loop operates during the calibration cycle with the same gains as it would if Canopus were in the field of view, thus avoiding the need to compensate for non-linearities in the AGC loop response.

In the normal operating mode, the intensity voltage resulting from whatever sources are within the sensor's field of view by-passes the A to D to A converter and appears directly at the upper and lower gate comparators. The signal is then compared with the reference voltages representing 1.5 times and 0.8 times Canopus intensity. If the signal is between these voltages, only the lower gate will be triggered and the acquisition mode will be initiated. If the signal is not between the gates, the acquisition logic will not be satisfied, and roll search will continue.

No interference between the acquisition and calibration functions is anticipated since it is expected that calibration would be done prior to but not during roll search.

A simplified timing diagram is presented in Figure 9. The calibration cycle is initiated by a single pulse from the spacecraft controller. The acquisition logic is "held" in its current state and the scan is deflected to the image of the calibration source. During this time, the A to D to A converter is reset to zero. After a preset period to allow the AGC loop transients to diminish, the digital counter is started and stores the calibration voltage digitally. The scan is then returned to its former operating position. After the AGC loop again is allowed to stabilize, the acquisition logic "hold" is removed and normal operation resumed.

CALIBRATION ERROR ANALYSIS

Ideally, it is desirable to set the acquisition gates as close to Canopus' intensity as possible. Practically, however, there is some minimum offset which is required to assure that Canopus will always fall between the gates in spite of the residual variations or errors which are present even with calibration. These errors and their combinations are discussed below.

a. Phototube Non-uniformities

CBS Laboratories, manufacturers of the RECONOTRON image dissector used in the Mariner IV sensor, has estimated that tubes could be selected whose total non-uniformity of the sensitivities of the photocathode and dynodes is less than $\pm$ 10% for a constant input image at least .005 inch in diameter. [6] This is the same as saying that, regardless of the star image's position on the photocathode, the tube output current will not fall to less than 82% of maximum. A tolerance of $\pm$ 10% is thus considered to be three times the standard deviation (3σ).

b. Calibration Source Variations

Variations in the light output of the calibration source clearly affect the gating levels since the source provides the basis for those levels. After considering several low level light sources, the Cerenkov radioisotope source was selected as being best suited for this application. A typical Cerenkov source consists of an isotope capsule (containing Strontium 90) and a quartz disk within which the light is produced by deceleration of emitted electrons. The half-life of Sr-90 is about 28 years, producing a first year decay in output of 2½%. [7] Browning effects in the quartz are thought to be negligible. Light from such a source can be "piped" to the RECONOTRON photocathode through a fiber optic bundle. Thus, a resonable 3σ value for source variation is $\pm$ 2%. Some data on the Cerenkov and electro-luminescent diode sources are given in References 7 & 8.

c. Calibration Electronics Error

The errors due to the calibration circuitry are produced by the following blocks shown in Figure 8:

1 - A to D to A Converter

2 - Amplifier

3 - Voltage Divider

4 - Upper Gate Comparator

5 - Lower Gate Comparator

The errors will fall in two categories -- a normal mode of operation with a temperature range of 0°F to 100°F, and a degraded mode of operation with a temperature range of -50°F to 150°F. The errors are all worst-case.

The A to D to A Converter contributes an error which is set by the number of bits. Minimum errors of 0.01% (which is much more accurate than required) are attainable. After examining the errors of other portions in the calibration loop, it was decided that the total electronics error should be less than $\pm$ 2.0% for a worst-case condition. The A to D to A converter was therefore designed to have an error less than $\pm$ 1.00% for the degraded mode. The error over the normal temperature range is about half that, or $\pm$ 0.50%.

The amplifier is a feedback-stabilized operational amplifier with open- and closed-loop gains on the order of 20,000 and 1.2, respectively. Total gain variations in the normal and degraded modes will be less than 0.1% and 0.2%, respectively.

The error contributed by the Voltage Divider at the A to D to A Converter output can be initially $\pm$ 0.1%, plus an additional $\pm$ 0.05% over a period of six months. Another $\pm$ 0.10% error is added for the degraded mode temperature range. Thus, the total Voltage Divider tolerance will be $\pm$ 0.15% for normal operation and 0.20% for degraded operation.

The comparators are integrated circuit differential amplifiers. The Upper Gate Comparator and Lower Gate Comparator are designed to have a maximum error of 0.1%.

Thus the total maximum error for the Automatic Calibration System electronics will be ± 2.0% for worst-case conditions as shown in Table 1.

Table 1

Function	Normal Temp. Error	Degraded Temp. Error
Amplifier	±0.10%	±0.20%
Voltage Divider	±0.15%	±0.20%
Upper Gate Comparator	± .10%	± .15%
Lower Gate Comparator		
Sum of Above Errors	±0.35%	±0.55%
A to D to A Converter	±0.50%	±1.00%
Total Electronic Error	± .85%	±1.55%
Maximum Error Allotted to Electronics		±2.00%

d. Absolute Calibration Error

In order to initially set the gating levels correctly, the sensor must be operated against a star simulator which accurately represents (to the sensor) Canopus intensity as seen outside the earth's atmosphere. Clearly, the accuracy of this simulation directly affects how closely Canopus' intensity will actually appear to its desired level between the acquisition gates.

Experience with OAO star simulator calibration and data from JPL on the Mariner IV program indicates that absolute star intensities can be simulated within 10% with the existing technology. Improvements in absolute intensity measurement are continuing. Of particular interest is a series of Aerobee rocket flights being conducted by NASA's Goddard Space Flight Center. (5) The star sensors on these vehicles are calibrated before launch and after recovery, and are currently being used to make fine corrections in Goddard's calibration standard. We assume that such continuing work during the next few years will reduce maximum calibration errors to ± 5% of the absolute star intensity.

e. Optical Attenuation

As the axis of an optical system is rotated away from a star, the illumination at the entrance aperture of the system diminishes, thus reducing the energy delivered to the phototube and changing the apparent brightness of the star. Meisenholder and Davis have stated that, without aperture compensation, the energy delivered by the Mariner IV sensor optics to the phototube would diminish to 66% of its on-axis value for a star 16° off-axis (the maximum offset of Canopus in cone angle is 14°). (4) Since a reduction to 66% is equivalent to ± 20% about a median value, it is clear that aperture compensation is required to implement in-flight calibration. Residual variations in apparent star brightness due to off-axis attenuation are assumed to be less than 5%. Aperture compensation used on Mariner IV was at least this good.

f. Signal Variation with Roll Error

Figure 10 reveals that in the roll error range between -2 degrees and +2 degrees, the telemetered intensity voltage is not constant, but varies due to combination of the Fourier series components in the tube output and the transmission characteristics of the 1 & 2 KH_z filter. The variations about a median value are found to be 10 to 15% of the value, and therefore must be corrected or avoided if in-flight calibration is to be useful. It can be seen, however, that over the range between -0.5 degree to +0.5 degree, the variation is very slight. Thus it is proposed that the "acquisition decision" be made while the star is within this 1 degree angular region during roll search, as indicated by a pair of threshold detectors on the output of the roll error channel. In this manner, these variations can be limited to 2% or less.

g. Error Summations

If the intensity of Canopus is designated as unity, Rigel's and Vega's intensity is approximately 0.6. Then the nominal setting of the lower acquisition gate should be 0.8 to simultaneous maximize the probability of accepting Canopus and rejecting Vega and Rigel. Thus, under worst-case tolerance conditions, the total variation in the gate must be less than ± 20% of Canopus' intensity. Referenced to the gate itself, the variation must be less than ± 25%. The "margin" (between the expected worst case and allowable errors) must be determined on the basis of the degree of confidence in the maximum individual variations. (See Tables 2 and 3)

Table 2

	Canopus Referenced	Lower Gate Referenced
Canopus	1.00	1.25
Lower Gate	0.80	1.00
Vega/Rigel	0.60	.75
Maximum Allowable Gate Variation	±0.20	±0.25

Table 3

Worst Case Tolerance Summation

Phototube non-uniformities	± .10
Calibration source variations	± .02
Calibration electronics error	± .02
Absolute calibration error	± .05
Optical attenuation	± .05
Signal variation with roll error	± .02
	± .26

A more likely value for the maximum variation can be found by considering the probability of occurrence of each. Since their behavior as a function of time will be well known, and because they can certainly combine at maximum value, the variations of the calibration source (which will deliberately appear too bright at the start of the mission) and of the optical attenuation (which will be maximum when the spacecraft is in or near the ecliptic meridianal plane containing Canopus) are added directly. To this "subtotal" can be added the root-sum-square combination of the other errors previously described, since these are uncorrelated and can be assumed normally distributed. This process is shown in Table 4.

Table 4

Statistical Combination of Calibration Errors

Error Source	Maximum Error	Variance	Error Summation
1. Calibration Source Variations	.02	----	.02
2. Optical Attenuation	.05	----	.05
3. Phototube Non-Uniformities	.10	.001109	
4. Calibration Electronics Error	.02(3σ)	.000045	
5. Absolute Calibration Error	.05(3σ)	.000289	
6. Signal Variation with Roll Error	.02(3σ)	.000045	
Sum of Variances =		.001488	
Standard Deviation (1σ) =	.0386		
3σ =	.116		.116
Max. Probable Error (3σ Equivalent) =			.186

EXPERIMENTAL SENSOR

An experimental model sensor is now being constructed to investigate the implementation of in-flight calibration. The design is a modified version of the JPL Mariner IV Canopus sensor. Although the optics are different, the model sensor includes a RECONOTRON image dissector and a functional duplicate of the Mariner IV sensor electronics. To these have been added a calibration light source, associated optics, and the new calibration circuits. As of this writing, the entire sensor has been assembled and checked out, and testing of the calibration function has begun.

CONCLUSIONS

A comparison of the allowable error in the lower gate level (25% as shown in Table 1) and the maximum probable error (18.6% as shown in Table 3) provides a significant margin for confidence that Canopus will always be acquired and other bright objects always rejected. Regardless of the confidence in the sensor's ability to uniquely acquire Canopus, however, acquisition will surely be verified by confirmation of the earthward orientation of the high gain antenna and by monitoring the observed field brightness as indicated by a telemetered AGC signal during all periods of roll search. This "map-matching" process was used on Mariner IV and found to be entirely satisfactory. The difference between Voyager and Mariner in this regard is that active ground control of Mariner's roll attitude was required, since the sensor could acquire many references, while on Voyager, only verification of Canopus acquisition should be required. Roll search command capability will undoubtedly be retained for redundancy, but will not be required in the normal operating mode.

The most significant conclusions resulting from the investigation are these:

(1) The stars Rigel and Vega are the most difficult objects to discriminate against since they are nearly as intense as Canopus and in addition, Vega is nearly opposite Canopus on the celestial sphere. While natural planetary satellites can pass through Canopus' intensity near encounter, the effect is brief, and the probability of falsely acquiring any of these is very small. The brightest portions of the Milky Way produce a sensor output well below Canopus' level.

(2) Non-uniform photocathode and multiplier sensitivities in the image dissector are the limiting factors in setting gate thresholds close to Canopus' intensity.

(3) Errors in absolute calibration (with respect to Canopus actual intensity) must be limited to a few hundredths of a magnitude.

(4) Of the several techniques that can be used to uniquely acquire Canopus, intensity discrimination with periodic in-flight calibration offers a high probability of success with minimum complexity.

67-585

(5) An attractive alternate method of star identification is that of using dual star sensors to simultaneously detect Canopus and Vega.

REFERENCES

1. Paper, "Mariner IV Canopus Sensor", G. W. Meisenholder and E. Davis (JPL), presented at the 16th International Astronautical Congress Symposium in Athens, Greece, Sept. 1965.

2. Special Report #110, Research in Space Science, Subject: Project CELESCOPE, An Astronomical Reconnaissance Satellite, Edited by R. J. Davis, Dec. 1962 (Cambridge).

3. PIR *4176-616, "Program: Voyager, Subject: False Targets for Canopus Tracker", H. E. Brown, 9 Dec. 1965.

4. "Integrated Starlight Over the Sky", F. E. Roach and L. R. Megill, Boulder Laboratories, NBS, Astrophysical Journal, Vol. 133, No. 1, Jan. 1961.

5. Personal Conversations with L. Draper, NASA's Goddard Space Flight Center, May 1967.

6. Letter from CBS Laboratories' R. C. Stow discussing photocathode variations, (Ref. CBS Proposal 66-233), 9 Nov. 1966.

7. Letter from 3M Company's T. N. Lahr discussing Cerenkov sources, 11 Nov. 1966.

8. Letter from Electro-Nuclear Laboratory's R. T. Chow discussing gallium phosphide diode light sources, 15 Nov. 1966.

9. PIR *4176-618, "Stellar Magnitude Determination", H. E. Brown, 13 Dec. 1965.

10. PIR *1200-VC-E300-VT-0075, "Canopus Gate Settings", H. E. Brown, 13 Dec. 1965.

11. PIR *41M8-013, "Design and Performance Data on the Mariner IV Canopus Sensor", G. B. Daly, 8 Sept. 1966.

12. PIR *41M8-EHS-003, "Low Intensity Light Sources for Star Tracker Calibration", E. H. Stockoff, 30 June 1966.

13. Basic specification for RECONOTRON Image Dissector.

14. Communications of the Lunar and Planetary Laboratory No. 63, University of Arizona.

15. Astrophysical Quantities, C. W. Allen, 1963 (London).

16. A Star Atlas and Reference Handbook, A. P. Norton and J. G. Inglis, 1957 (London).

* Internal correspondence of the General Electric Company, Missile and Space Division.

STAR NAME	POSITION				SPECTRUM		INTENSITY	
	RIGHT ASCNSION		DECLINATION		CLASS	COLOR	MAGNITUDE	
	HRS.	MIN.	DEG.	MIN.			VISUAL (1)	S-4(2)
SIRIUS	6	41	-16	35	A1 V	BLUISH-WHITE	-1.46	-1.46
CANOPUS	6	22	-52	38	F0 Ib	YELLOW-WHITE	-0.71	-0.56
RIGEL	5	10	- 8	19	B8 Ia	BLUISH	+0.15	-0.03
VEGA	18	34	+38	41	A0 V	BLUISH-WHITE	+0.02	+0.02
AGENA	13	57	-59	53	B1 II	BLUISH	+0.61	+0.13
ACHERNAR	1	34	-57	45	B5 IV	BLUISH	+0.49	+0.17

(1) H.L. JOHNSON, COMM. OF THE LUNAR AND PLANETARY LAB. NO.63, UNIV. OF ARIZ.

(2) AEROBEE AND OAO PROJECTS, NASA GODDARD

FIGURE 1: CHARACTERISTICS OF THE SIX BRIGHTEST (S-4) STARS

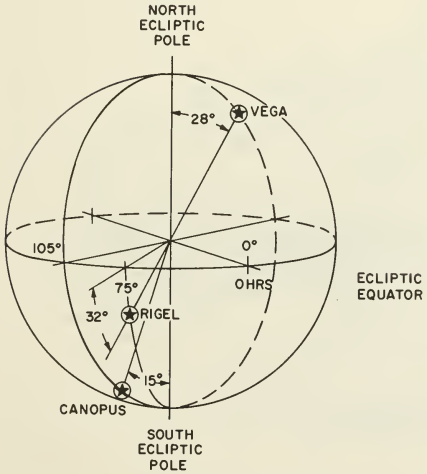


FIGURE 2: CELESTIAL GEOMETRY OF CANOPUS, RIGEL AND VEGA

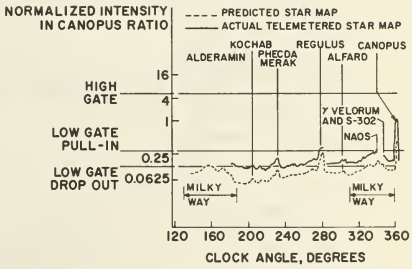


FIGURE 3: PREDICTED VS. ACTUAL SENSOR SIGNALS FOR THE FIRST ROLL SEARCH OF MARINER IV (From Meisenholder & Davis, JPL)

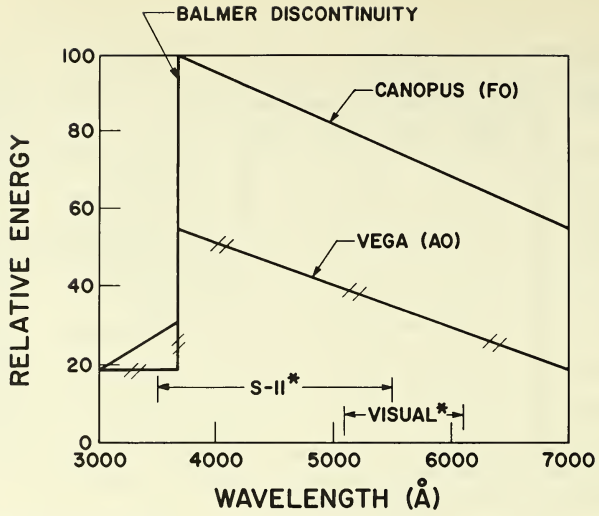


FIGURE 4: SIMPLIFIED SPECTRAL IRRADIANCES CANOPUS AND VEGA
(From Report on Project CELESCOPE, R.J. Davis, 1962, Cambridge)

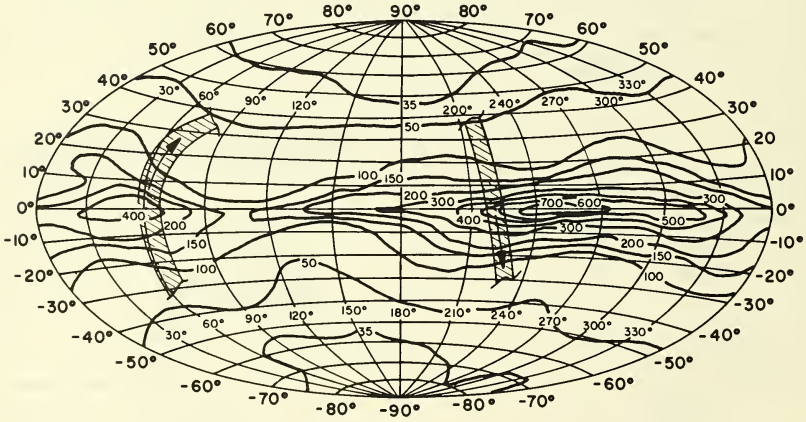


FIGURE 5: TOTAL INTEGRATED STARLIGHT IN NUMBER OF TENTH-MAGNITUDE (VISUAL) STARS PER SQUARE DEGREE IN GALACTIC CO-ORDINATES
(From Reach & Megill, Boulder Labs, NBS, 1960)

BRIGHTEST STARS	S-4 INTENSITY	ACQUISITION GATES	
		MARINER IV	AUTOMATIC SENSOR
SIRIUS	2.30	8.00	1.50
CANOPUS	1.00		⇕
RIGEL / VEGA	0.62 / 0.58		0.80
AGENA	0.54		
ACHERNAR	0.51		
REGIL KENTAURUS	0.47		
CAPELLA, PROCYON, ARCTURUS	0.32	0.12-0.30	

FIGURE 6: STAR INTENSITIES AND GATING LEVELS
(ALL VALUES NORMALIZED TO CANOPUS)

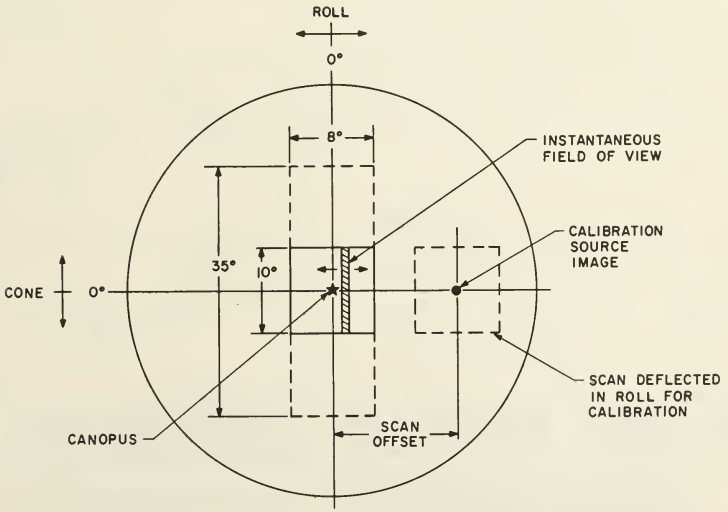


FIGURE 7: SCAN PATTERNS ON PHOTOCATHODE

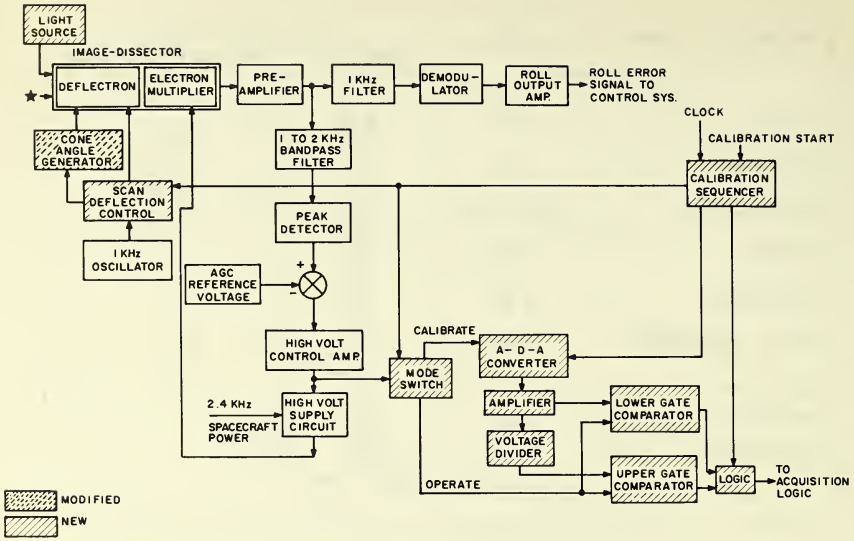


FIGURE 8: AUTOMATIC CANOPUS SENSOR BLOCK DIAGRAM
(Original Mariner IV Sensor Diagram From Meisenholder and Davis, JPL)

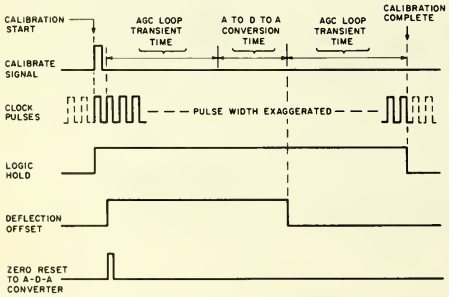


FIGURE 9: CALIBRATION SEQUENCER TIMING DIAGRAM

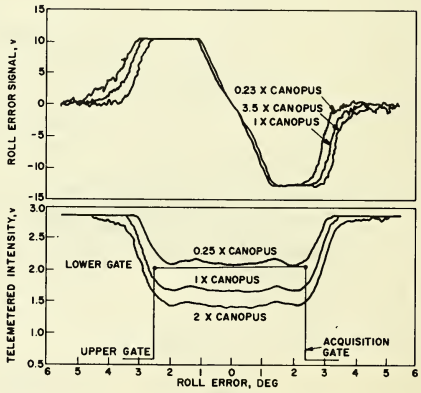


FIGURE 10: ROLL OUTPUT VOLTAGE, ACQUISITION GATE VOLTAGE, AND APPARENT STAR BRIGHTNESS VS STAR ROLL POSITION (From Meisenholder and Davis, JPL)

No. 67-586

CALIBRATION AND SPACEBORNE PERFORMANCE OF THE SURVEYOR SPACECRAFT STAR SENSOR

by

JACK C. LANSING, JR.
Santa Barbara Research Center
Goleta, California

AIAA Paper
No. 67-586

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

-- NOTES --

CALIBRATION AND SPACEBORNE PERFORMANCE OF THE SURVEYOR SPACECRAFT STAR SENSOR*

Jack C. Lansing, Jr.
Member of the Technical Staff
Systems Engineering Section
Santa Barbara Research Center
Goleta, California

ABSTRACT

The Surveyor spacecraft navigational star sensor identifies Canopus by intensity using the sun as a reference. The calibration of the sensor through the atmosphere is described, with statistical analysis of the calibration results showing the practicality of low altitude calibration. As the first Surveyor rolled to find Canopus, its star sensor scanned twelve recognizable stars. A method of analyzing the star intensity signals is described and applied to seven of the stars, partially verifying calculated values for these signals. The intensity of the other stars aids in an estimate of the sensor calibration in terms of Canopus, which is found to give a signal 35% higher than expected. In the absence of stars, the star intensity channel signal was increased 23% over the minimum expected value either by increased gain or stray light. The method of stray light testing of the sensor is described. Performance in the second and third Surveyors is described briefly and generally verifies the conclusions drawn in studying the first mission.

INTRODUCTION

The Surveyors are accurately guided to landing locations on the moon with the aid of a reference furnished by the star Canopus, prime star of the ancient constellation Argo, ship of the Argonauts; thus, the star is a link between the mythical voyage for the Golden Fleece and our modern voyage for scientific discovery.

This star, ideally located for space navigation at nearly a right angle to the sun, has another very useful attribute. Canopus, second brightest of the stars, differs by more than one-half magnitude from the brightest and third brightest stars. This difference permits identification by the Canopus sensor developed for the Surveyor spacecraft. The sensor uses the sun as a comparison standard to perform the identification.

The problem discussed in this paper is the calibration of the sensor to recognize the Canopus-sun brightness ratio in space, on the basis of measurements through the earth's atmosphere, including the additional restriction that the two sources are not usable simultaneously. The sensor's performance on the historic trip of Surveyor I is analyzed to verify the calibration. Pertinent information from the later Surveyor flights is added to bring the analysis up to date.

The design of the sensor has been previously described, [1] and thus will only be covered sufficiently here for understanding the calibration and performance. Figure 1 is a photograph of the sensor, and Figure 2 is a block diagram of its functioning. The starlight and sunlight, the latter much attenuated, are chopped at different rates and detected by a photomultiplier tube. Electronic filters separate the star and sun signals, and the sun signal operates an automatic gain control which maintains the sun signal nearly constant at the filter output. As a result, the star signal amplitude at its filter output is proportional to the ratio of star to sun. The sensor outputs telemetered to

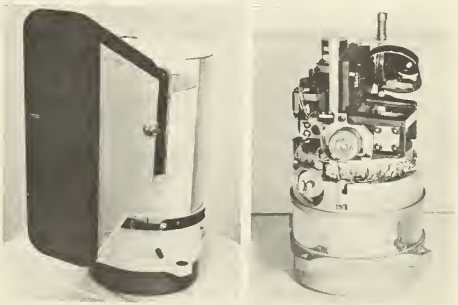


Figure 1. Canopus Sensor Exterior View and View with Cover Removed

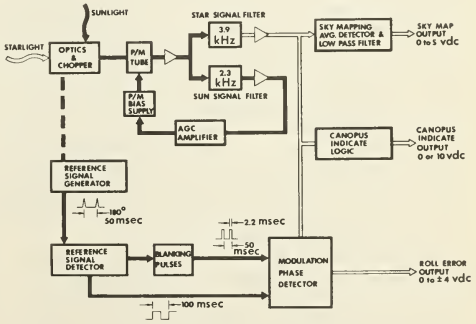


Figure 2. Canopus Sensor Functional Block Diagram

*This work was supported by a Hughes Aircraft Company contract, and Jet Propulsion Laboratory, California Institute of Technology, Contract JPL-90056.

ground stations are the star mapping or intensity signal, star roll error signal, and Canopus indicate signal. At the control center the first two signals are reproduced on a chart as received, and the last is displayed as a digital indication since it is a two-state signal. Table 1 lists some characteristics of the sensor.

Table 1. Canopus Sensor Characteristics

Telescope Aperture	1 inch
Instantaneous Field of View	4° in roll x 5°
Scan Angle	±2°
Scan Rate	10 Hz
Roll Error Output Linear Region	±2°
Roll Error Output Accuracy	±0.1°
Starlight Chopping Rate	3.9 kHz
Sunlight Chopping Rate	2.3 kHz
Photoresponse	S-4
Identification Threshold Factors	0.67x, 1.50x Canopus

In order to test and calibrate the sensor, sources of radiation equivalent to Canopus and the sun outside the atmosphere were developed. These are embodied in a sun-star simulator which derives the two sources from a single lamp, thus minimizing relative variations between the simulated sun and star. To avoid the problems of producing actual sun level radiation with good stability, a removable attenuating filter is used in the sensor sun channel. The filter is removed when using simulator input, and inserted for the flight.

SIMULATOR CALIBRATION

The calibration of the simulator is performed by using a Canopus Sensor to compare the radiation from a star with that from the simulator star channel, and the radiation from the sun through the filter with that from the simulator sun channel with the filter removed. In each case (sun or star), the ratio of signal voltage from the actual object to that from the simulator is obtained at intervals as the object moves through the sky. Since the effective quantity of atmosphere through which the radiation of sun or star travels is proportional to the secant of the zenith angle (sec Z), if Z is not too large, the ratios may be graphed against sec Z and extrapolated to sec Z = 0 to arrive at the value of the ratio outside the atmosphere. [2]

Since Canopus is so far south as to be unusable for calibration in the U. S. A., Sirius was chosen for this purpose because of its color similarity to Canopus, its location, and its high intensity. The calculated Sirius-Canopus intensity relationship was verified using Canopus sensors and sun-star simulators in field measurements near Santiago, Chile.

The method of calibration used is subject to error from changes in the atmosphere during the time of calibration which, by changing the slope of the graphed line, changes the extrapolated ratio. Ideally the calibration would be performed at a high elevation astronomical observatory, but questions of availability and economics led to work being performed first at a medium altitude field location in the Santa Barbara area and later at a 2600-foot elevation, Tucson, Arizona, facility of Hughes Aircraft Company, to gain the advantages of more clear days per year and the use of an observatory building at that facility.

Some samples of the signal ratio versus sec Z data obtained are shown in Figure 3. Following the practice of astronomy, the ratio is expressed as the stellar magnitude difference, defined as

$$\text{mag} = -2.5 \log_{10} (V/V_o)$$

where mag = magnitude
V_o = circuit voltage generated by the reference irradiance
V = circuit voltage generated by the sun or star irradiance.

The lines plotted have been fitted to the points by the method of least squares, and have the equations

0.12 mag + 0.27 mag sec Z (3/3/66)
0.02 mag + 0.27 mag sec Z (3/4/66)

The scatter of the data points about the line is a little small for the 4 March data, in comparison to the general run of data obtained.

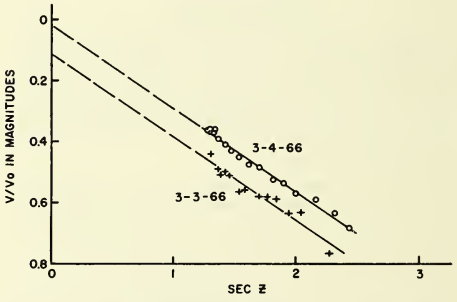


Figure 3. Sample Extrapolations of Sun/Simulator Sun vs Secant Z or Atmospheric Thicknesses

STATISTICAL ANALYSIS OF CALIBRATION

Of the sec Z = 0 intercepts which have been derived from the calibration measurements, the largest groups obtained under unchanged conditions (except those dictated by the atmosphere) were selected. While these are few for the purpose, they have been analyzed statistically to furnish an indication of the validity of the data. The first group of measurements was made on the sun over a six-week period and resulted in eleven intercepts. (See Table 2.)

Table 2. Equations of Lines Fitted to Sun Data Runs

Date	Least Squares Line		σ* (mag)
	Intercept (mag)	Slope (mag/sec Z)	
2/28	0.18	0.25	0.009
2/28	0.09	0.26	0.015
3/3	0.12	0.27	0.020
3/4	0.02	0.27	0.011
3/7	0.00	0.32	0.012
3/16	0.14	0.32	0.010
3/28	0.13	0.37	0.014
3/29	0.04	0.40	0.017
3/30	0.11	0.33	0.014
3/31	0.07	0.35	0.015
4/7	0.13	0.26	0.022

The arithmetic mean of the intercepts is 0.09 magnitude, which indicates that the simulator sun is bright by this amount with respect to the real sun seen through the sun filter. The difference was corrected by making a new filter of increased transmission.

Graphs of the intercepts versus slopes of the lines fitted to the data, versus the rms departure of the data points from the lines, and versus time, show no apparent correlation of the intercept with any of these parameters. See Figure 4. On the basis of these results, it might be assumed that the intercepts are randomly distributed values. If so, the rms error of the arithmetic mean of the intercepts will be[3]

$$\mu_{N(SUN)} = \sigma_i \, (N - 1)^{-\frac{1}{2}}$$

where σ_i = the rms value of the differences of the intercepts from the mean intercept
= 0.053 mag
N = number of intercepts = 11
or substituting, $\mu_{N(SUN)}$ = 0.017 mag.

The second group of measurements, on Sirius, have resulted in the data of Table 3.

Table 3. Equations of Lines Fitted to Star Data Runs

Date	Least Squares Line		σ* (mag)
	Intercept (mag)	Slope (mag/sec Z)	
1/23	-0.97	0.27	---
2/11	-0.89	0.30	0.014
2/12	-0.94	0.31	0.026
2/13	-0.89	0.29	0.017
2/19	-0.79	0.36	0.014
2/21	-0.87	0.35	0.019
2/27	-0.94	0.38	0.027
3/3	-0.86	0.29	0.025
3/4	-0.94	0.29	0.014

*σ = standard deviation of the data points from line.

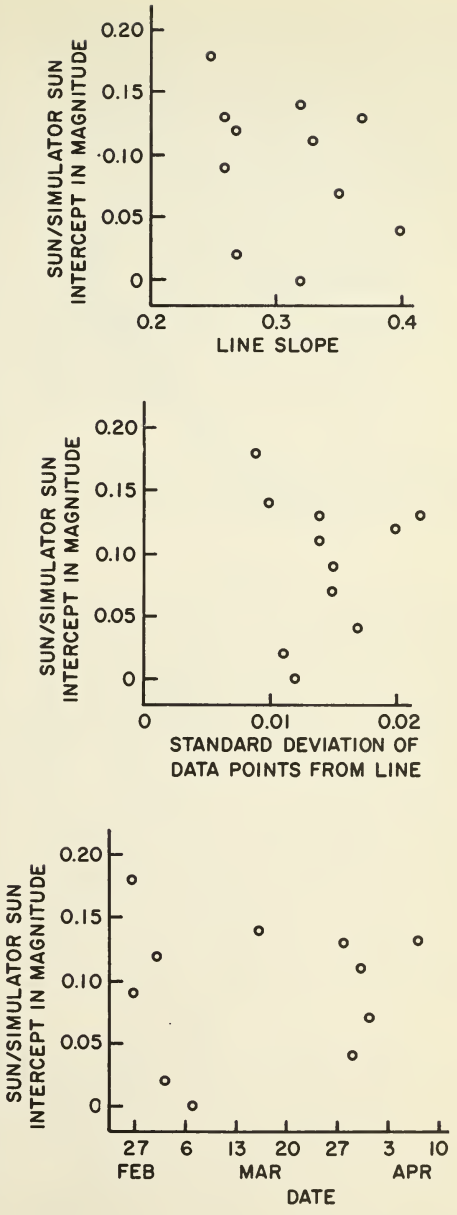


Figure 4. Sun/Simulator Sun Zero Atmosphere Intercepts as a Function of Several Parameters

67-586

The mean of the intercepts is -0.90 magnitude. Since this is exactly the S-4 magnitude difference between Sirius and Canopus, the simulator star is at Canopus intensity. In the same manner as for the sun, we may calculate the rms error of the mean for the star as $\mu_{N(STAR)} = 0.018$ mag.

To estimate the maximum error in Canopus to sun ratio to be expected on the basis of the statistics, the following equation is used

$$E_{max} = nb\,[\mu_{N(SUN)}^2 + \mu_{N(STAR)}^2]^{\frac{1}{2}}$$

where n = a factor adjusted to take in the desired portion of the assumed gaussian error curve; for 99.7% probability of including the error, n = 3 is used.

b = a factor to allow for the uncertainty in the μ_N values due to the small sample. Braddick[3] states that with ten observations there is about a 97% chance that the calculated rms value is within 50% of the correct rms value. b = 1.5 is used.

Substituting, $E_{max} = 0.11$ mag. This value is in reasonable agreement with the estimate of ±10% which had been made before the measurements which were analyzed.

STAR MAP

Figure 5 shows a reproduction of the first star intensity data from Surveyor I obtained when the spacecraft was rolled to produce a "map" enabling identification of Canopus. The star chart in the figure shows the path covered by the sensor field of view during the spacecraft roll, and the constellations crossed by this path. The star intensity waveform was continuous as received, but is broken at the equator in the figure to align it with the star chart.

Before the flight, a series of simulations of expected star maps, based on calculated star amplitudes, was computer generated for each possible launch day. A comparison of sensor output during one revolution with the simulated star map was sufficient to allow positive identification of Canopus so that commands could be given, first to stop at the approximate Canopus location, and then to lock onto the sensor's error signal. The dip in the star intensity waveform after the roll stop command is caused by an overshoot in stopping the roll.

STAR MAGNITUDE CALCULATION

The method of calculating the star magnitudes for the simulated maps is described in references 4 and 5. To briefly summarize this method, data are available on a few stars in the form of "monochromatic magnitudes" which are employed to find a color correction which can be applied to data available on many stars (V and B-V data) to give an S-4 photoresponse. These monochromatic magnitudes are narrow band (10Å), absolute flux measurements at twelve wavelengths through the S-4 band of approximately 3000Å width, while B and V are relative measurements in two bands of about 1500Å width each. The Canopus sensor uses an S-4 photoresponse, although it is modified, principally by the lens transmission. As reported in reference 4, the results of this method have been verified for stars of restricted types with spectral intensity characteristics similar to Canopus.

Astronomers classify the stars on the basis of their spectral absorption lines into the main classes O, B, A, F, G, K, and M, with ten Arabic numeral subdivisions. They are also classified according to their luminosity, designated by Roman numerals. Most of the monochromatic magnitude data are for type V stars. Canopus, type F0Ia, closely corresponds in color to a type A6V for S-4. The derivation made in reference 4 was based on data from star types B3V, A0V, A5V, F0V, F2V, G0V, G4V, and K5III. Data taken at SBRC on stars of types A0V, A1V, A2Ia, A7V, F5V, and F5Ib fell within a probable error of ±0.03 mag of the derived relation.

The color correction curves, reference 4, Figure 1, and reference 5, Figure 5, while starting from the same monochromatic magnitude data, show a somewhat different slope. Two differences were noted between the references. The S-4 photosurface spectral responses used are different, and Hyde[4] adjusted the assumed "V" filter response as a result of cross-check calculations, to obtain agreement among the various data available.

Fortunately, the stars of the roll map are mostly of types with color corrections near that of Canopus, thus minimizing errors from this source. Also, since the colors of most of these stars are similar to Canopus, the color characteristics of the optical elements of the sensor which the starlight traverses will have small effect.

STAR SIGNAL AMPLITUDES

The waveshape of the star mapping output during roll of the spacecraft is a truncated triangle. It takes this form partly because of the modulation used. The 4° field of view of the sensor is scanned linearly ±2° (and an additional nonlinear amount during turn-around). A centered star produces signal during the entire linear scan, while an off-center star produces signal for only part of the scan. The roll error signal is derived from the proportion of the time the star signal is present. Since the scan cannot be made linear in the region of turn-around, the turn-around time or overscan period is rejected in the logic processing. However, the mapping output is generated by linear detection of the signal before logic processing, and thus includes the overscan period, and resultant gap in star signal.

Sketches of typical star signal envelopes at the output of the 3.9-kHz filter are shown in Figure 6, from which it can be seen that during a roll the amplitude output will increase up to the "small roll error angle case," then will remain the same through the central portion of the field, then decrease. The star mapping signal waveform produced while rolling will show some rounding caused by the 1-second RC filter on the output of the circuit.

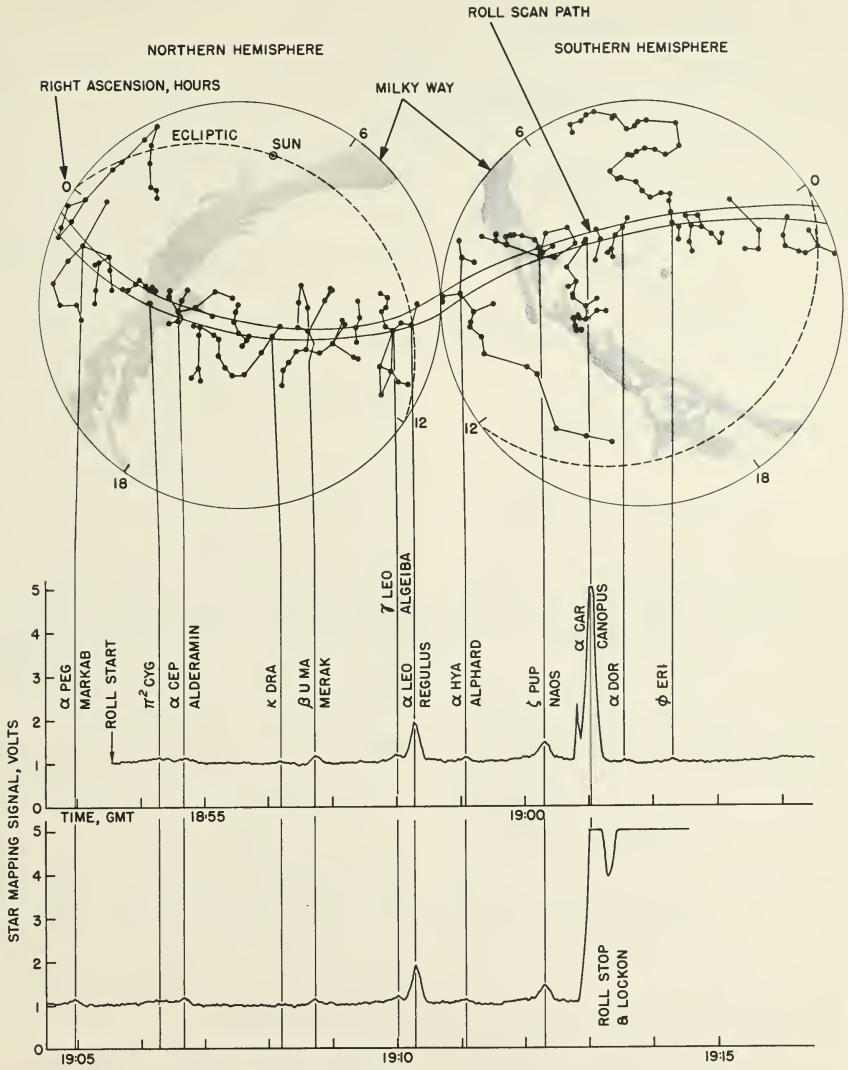


Figure 5. Roll Path of the Sensor Field of View and Star Mapping Signal

67-586

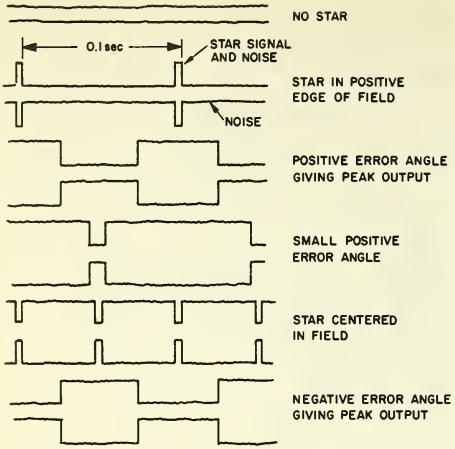


Figure 6. Typical Star Signal Waveform Envelopes at the 3.9-kHz Filter Output of the Sensor

The star amplitude which would be present without noise was obtained by, first, extrapolating the straight portions of the amplitude output waveform produced during a roll to intersection to find the amplitude output which would be produced, were the gaps in the centered waveform to be filled in, and were the filter not to round the detected waveform. This is the linear detected value of the portion of the signal labelled "Star Signal + Noise" in Figure 6.

For weaker stars, with the triangular roll waveform distorted, a transparent template with a set of triangles of different amplitude, but the fixed base width characteristic of a constant roll rate, was used to estimate the star signal amplitude. In any case, star strong or weak, the noise amplitude is available when the star is clear of the field. Figure 7 shows a plot of the data from both passes on the star Merak, with the triangle as fitted by the above method drawn in.

Goldman[6] gives an expression for linear detector output as a function of rms signal-to-noise ratio. The dc component of the output is

$$V_{DC} = \frac{\alpha R}{2A\pi^{\frac{1}{2}}} \exp(-A^2K^2/2) \times [(1 + A^2K^2)I_0(A^2K^2/2) + A^2K^2I_1(A^2K^2/2)]$$

where α = detector conductance when conducting
R = detector load resistance
$A = 1/[2^{\frac{1}{2}} \times (\text{rms noise input})]$
$K = 2^{\frac{1}{2}}$ (rms sine wave input)

and I_0 and I_1 are modified Bessel functions of the first kind and of zero order and first order, respectively.

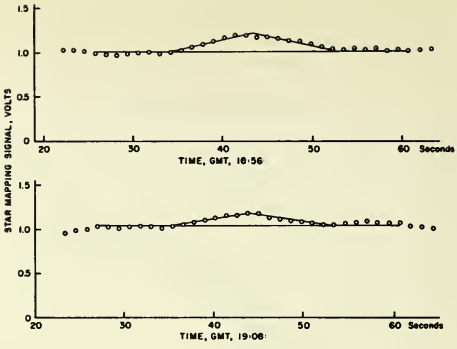


Figure 7. Triangles Estimated to Fit Signal from the Star Merak

For pure noise this equation reduces to

$$V_{DCN} = \frac{\alpha R}{2A\pi^{\frac{1}{2}}}$$

so that the ratio of linear detector outputs with and without star is

$$V_{DC}/V_{DCN} = \exp(-A^2K^2/2) \times [(1+A^2K^2)I_0(A^2K^2/2)+A^2K^2I_1(A^2K^2/2)]$$

This expression was computed and tabulated over a range of values of AK, which is the ratio of rms star signal to rms noise, to facilitate finding values of AK from the measured values of V_{DC} and V_{DCN}. With this we can calculate the star output which would be obtained in the absence of noise

$$V_{DCS} = AK \times V_{DCN} (N/N_{av}) (S_{av}/S)$$

where N/N_{av} = (rms value/full-wave rectified average value) for gaussian noise
$= (\pi/2)^{\frac{1}{2}}$
S_{av}/S = (full-wave rectified average value/rms value) for sine wave
$= 2 \times 2^{\frac{1}{2}}/\pi$

For comparison purposes the star output voltages are converted to magnitudes, using Regulus as the reference, since Canopus is probably limiting.

Table 4 lists the stars which appeared to show some signal in the roll map, their spectral classes, their magnitudes as calculated before the mission, and their magnitudes found in the analysis of the roll map described above.

ESTIMATE OF SENSOR CALIBRATION

In this roll map Canopus signal is high enough to be in the electronically limited region of response. In a recent test of one sensor it was found that signals in the range of factors of approximately 1.2 to 8 times nominal Canopus signal are limiting and produce roll mapping signals which vary little in

Table 4. Stars Appearing in the First Surveyor Roll Map

Star Name	Spectral Class	S-4 Magnitude	
		Calculated	Measured*
α Car Canopus	F0Ia	-0.57	-0.07
α Dor	A0p	+3.62	---
ϕ Eri	B8V	+3.43	---
α Peg Markab	B9V	+2.46	+2.34
π^2 Cyg	B3	+4.27	---
α Cep Alderamin	A7IV, V	+2.62	+2.47
$^{\kappa}$ Dra	B5p	+3.93	---
β UMa Merak	A1V	+2.34	+2.30
γ^1 Leo Algeiba	K0pIII	+3.58	+2.32
α Leo Regulus	B7V	+1.22	+1.22
α Hya Alphard	K3III	+3.31	---
ζ Pup Naos	O5	+2.03	+1.79

*Measurements are referenced to Regulus.

appearances, even though signals in the lower part of this range are effective in the "indicate" circuit. Canopus signal was expected to be high, as a result of a precautionary gain increase in the sensor, which was necessary because of a possible problem discovered late in tests of the Surveyor. The problem was that should the Canopus sensor temperature approach the maximum thought to be possible during the flight, condensation could form on the inside of the window through which the star telescope looks and which is cooled by radiation to space. An allowance was made, in the first Surveyor, for this possibility by increasing the star signal amplifier gain 50%, such that with the maximum expected condensation, Canopus signal would yet be strong enough for guidance. On later flights, a change in thermal configuration will raise the window temperature to avoid the problem.

To estimate the sensor calibration, we assume the next brightest star, Regulus, to have the sensor effective magnitude in space which was calculated, and find that the other stars take the values shown in Table 4. Taking the stars in descending order of intensity, Naos was 0.24 mag brighter to the sensor than calculated, but is a type O star, thus outside the range of stars for which the calculation is known to be valid. Merak is very close to its calculated value. Algeiba, at 1.26 mag brighter than calculated, has a large color difference from Canopus, but is still rather surprising. η Leo and ζ Leo can be in the sensor field when Algeiba is near center, and may be adding to the signal. Markab and Alderamin are respectively 0.12 mag and 0.15 mag brighter than calculated, but here the signal-to-noise ratio is becoming low, and the remainder of the stars are not quantitatively readable.

The Canopus value as measured is 0.50 mag low, by this comparison to the other stars. This difference is assumed to be caused by the electronic limiting mentioned above.

The test data on this sensor before flight show that nominal Canopus mapping signal should be 4.45 volts. Were this value extrapolated to a triangular peak as were the actual roll map values, it would be 4.65 volts, which is 0.26 mag lower than the measured value in flight. Had the limiting not occurred, the true value of Canopus signal would have been -0.26 mag -0.50 mag = -0.76 mag compared to nominal Canopus signal. From this we deduct the deliberate gain increase for possible window fogging of 1.5× or -0.44 mag to find the amount by which the signal exceeded the expected amount, with the result of -0.33 mag. The upper threshold of the "Canopus indicate" circuit is at a factor of 1.5 times nominal Canopus, with test limits at ±10%, such that the indication is positive at 1.35 times nominal and negative at 1.65 times nominal. 1.35 times nominal corresponds to -0.33 mag, so that Canopus would have just been within the automatic identification region, without the precautionary gain increase.

SOURCES OF ERROR

Of the 35% error observed between Canopus signal during this mission and the expected signal, a maximum of 10% each is assigned to:

1. Changes which environmental stresses can produce in Canopus sensor characteristics, such as circuit gains, photomultiplier differential sensitivity to the sun and Canopus, and reflectivity or transmissivity of optical surfaces.
2. Errors in calibrating the sun-star simulator with light that passes through the earth's atmosphere.

In addition to these, there is the possibility of dirt collecting on the sun or star entry windows between the time of the last inspection and operation in space. Comparing the geometry of the two windows, the sun entry appears to be more susceptible to such a possibility, as it faces either upward before launch or toward solar radiation pressure after launch. If this window became dirty, star signal would increase. From observation of the unchanged mapping signal levels during the 58 hours between the first and last signal available, it can be concluded that there was no such collection during most of the flight.

The remainder of the observed error may be accounted for principally by the variation among sensors of the ratios of star and sun to simulator star and sun, which have been measured for some units. This variation arises from optical focus effects. There are additional errors known to be present such as transfer of calibration from the Tucson facility through portable simulators, and the precision with which the sun filters can be adjusted. These additional errors amount to a percent or two each and are not major contributors to the total error.

SIGNALS NOT ON COMPUTED MAP

Besides the stars appearing in the roll map, there are also apparent some other departures from the fairly constant low level of signal present during much of the roll map. Just before the first appearance of Canopus there is a sharp upward

spike, which cannot be due to a star or other distant stationary object, as it appears and disappears within 6 seconds. Figure 8 is an expanded plot of roll error and mapping signals during the time of this spike. The spacecraft is rolling at a slow and uniform rate of 0.5 degree/second, which with the roll field of view of greater than 8 degrees requires longer than 16 seconds to pass a star. The spike could have been caused by an object moving through the field of view, such as a sun-illuminated dust particle thrown loose from the spacecraft.

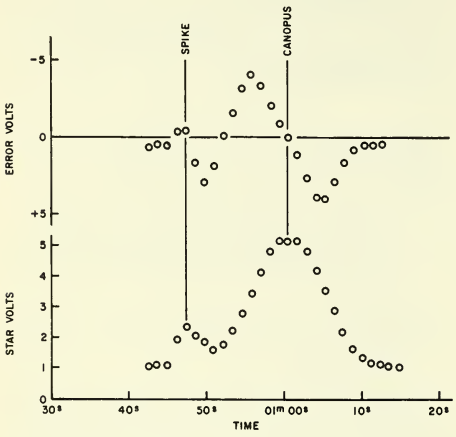


Figure 8. Signals at the Time of the Spike

An interesting analysis of a similar anomaly at the time of pressurization of the helium system just prior to the midcourse maneuver appears on page 167 of reference 7. After the detonation of a squib actuating a valve which passes helium into the vernier rocket engine system, signals appear in the Canopus sensor outputs. The analysis shows that the signals can be explained by a cloud of particles from the valve passing through the sensor field of view, and single particles from the line which is being pressurized, passing through the field at a later time. The elapsed time from squib firing to field entry and the time to cross the field corresponds to uniform velocity for the single particles, traveling from reasonable assumed spacecraft sources.

Other signal changes noticed are increases in the signal level while scanning across the Milky Way and the plane of the ecliptic. During the Santiago tests, a sensor was pointed at the brightest part of the Milky Way without producing a detectable signal, but two effects would increase the response in space: lack of atmospheric attenuation would increase signal, and absence of sky glow would increase the contrast between the Milky Way and the darker areas of space. The signal in the plane of the ecliptic may be from "zodiacal light" which is believed to be sunlight reflected from innumerable small bodies scattered in the ecliptic region. [8]

The level of map output during the apparently dark portion of the roll is a function of star channel gain and of any stray light entering the sensor. A gain increase of 50% from nominal was found to increase the dark noise by 30% from nominal in a laboratory Canopus sensor experiment. As was discussed above, the signal from Canopus was -0.76 mag from nominal, or a factor of 2.01 too bright. This can represent either a high gain due to low sun intensity, which controls the gain, or a miscalibration of the star channel such that the effective star intensity was greater than nominal. Of course, it is expected that the sun is at least a factor of 1.5 low because of the darker sun filter used. Were this the only departure of sun intensity from nominal, the output in the dark would be expected to be 1.3 x 0.63 volt = 0.82 volt; the flight "dark" level was approximately 1.01 volts, or 23% over the minimum expected value.

STRAY LIGHT EFFECT

In development of a baffle to protect the sensor from sunlight or sunlight reflections, quite small quantities of light were found to cause measurable effects on sensor outputs. An experimental setup to evaluate baffle performance without introducing uncontrolled stray light was a problem. It was conveniently solved when Jet Propulsion Laboratory made available a facility consisting of a large room with very low reflectance walls and floor, with a coelostat mirror arrangement to direct sunlight downward in the center of the room. A Canopus sensor was mounted on a Surveyor spacecraft in the column of sunlight, with a small lamp mounted on the wall to represent Canopus. The sunbeam was not large enough to cover all those portions of the spacecraft which could reflect interfering sunlight into the sensor sunshade opening, so as each portion was positioned in the beam, a measurement of illumination at the sunshade opening was made. The total of these measurements (increased to allow for atmospheric absorption) then was simulated by a spotlight replacing the part of the spacecraft most effective in causing interference, which was one of the landing feet. Even though the beam from the spotlight was restricted by baffles to just that portion which would enter the sensor sunshade, it was found that the light scattered by the air and suspended particles contributed to interfering signals, which in space would be due only to reflections within the sunshade. This effect was cancelled out of the results by comparing measurements under the above conditions with measurements made with the spotlight beam passing just in front of the sunshade, so that air and particle scattering was the only source of interference.

From these tests it was concluded that additional intensity channel output of the order of 0.2 volt due to spacecraft reflections could be expected, but that this quantity differs from sensor to sensor and is nonlinear with light level.

The conclusion with regard to the space flight can be only that there are qualitative explanations for the increased dark level.

Another possible source considered for the high no-star intensity signal is Van Allen Belt radiation noise in the photomultiplier. During the acquisition roll the distance from the earth was approximately 6.5×10^4 km, which is in or near the upper fringes of the belt.[9] However, at midcourse roll (1.6×10^5 km from earth), the no-star intensity appears unchanged, thus the belt evidently did not affect the sensor.

CANOPUS INDICATE OUTPUT

The Canopus indicate circuit output should be 10 volts for a signal between 0.67 and 1.5 times nominal Canopus signal, and zero volts outside this region. The large Canopus signal (or high sensor gain) caused the indicate signal to remain off while Canopus was present. The indicate output while rolling showed the effect of the increased dark level, switching on intermittently on the order of 10% of the time during the acquisition, midcourse, and preterminal rolls, with the on time increasing when star signals were visible. The indicate signal does not change appearance greatly during any of the times observed, while star signals are absent.

LATER MISSIONS

Surveyor 2 had been roll stabilized by the Canopus sensor for some hours before the spacecraft was lost when one of its three vernier engines would not fire.

The star signal amplifier gain of that sensor had been set 17% high against the window condensation possibility previously described. Once again Canopus signal limited in the electronics. Analysis of the roll map from this mission led to an estimate that Canopus signal exceeded the expected amount by -0.28 mag, after discounting for the 17% gain increase.

After Surveyor I and the calibration work described earlier it was discovered that the prototype sensor being used for calibrating the simulator did not respond the same as flight model sensors because of focus problems. A change was made to one of the flight model sensors which had been used in type approval tests, to be used as the calibrating standard. The recalibration indicated that with the old calibration, Canopus should appear -0.22 mag too bright, which is in rough agreement with the estimates from flight data.

The Canopus sensor for Surveyor III included the thermal configuration changes necessary to avoid the possibility of condensation on the window, so that no compensating increase was necessary. In addition, the results of the recalibration with the flight model sensor were available, so the gain of the star signal amplifier was reduced by +0.24 mag to allow for that change. The flight data from Surveyor III show Canopus signal exceeding the expected amount by only -0.08 mag. Furthermore, in this mission the sensor acquired Canopus in the automatic mode, which means that it performed the automatic identification of Canopus for the first time in space.

To make an intercomparison of all missions on the basis of the recalibration mentioned above, estimated Canopus signal in flight exceeds the expected amount as follows: Surveyor I, -0.11 mag; Surveyor 2, -0.06 mag; and Surveyor III, -0.08 mag. All of these values are well within the Canopus identification thresholds.

Additional items of information from the second and third missions are the following.

The level of map output during the apparently dark part of the roll remained high in these missions. In summary, the amount by which the flight "dark" level exceeded the level when tested in the dark was: Surveyor I, 60%; Surveyor 2, 44%; and Surveyor III, 28%. From these results we conclude that the decreased sun levels of I and II were causing some of the noise increase, but III shows that a contribution is being made by stray light.

The roll map of Surveyor III shows unidentified objects resembling stars but appearing only in the first revolution. The waveforms differ from those of stars only in being slightly irregular, as if brightness or position were changing a little while the objects are in the field.

CONCLUSION

The method of calibration of the sensor produces results of sufficient accuracy to allow identification of Canopus through amplitude comparison to the sun. The measured relative magnitudes of the stars as compared to each other are in good agreement with calculated values, for stars of spectral types B7V, B9V, A1V, and A7IV, V, but not for type O.

REFERENCES

1. Sochel, A.H. and Peterson, E.W., "The Surveyor Canopus Sensor," Proc. IRIS, Vol. 8, No. 3, pp 5-9 (August 1963).
2. Hardie, R.H., "Photoelectric Reductions," Astronomical Techniques, edited by W.A. Hiltner (University of Chicago Press, Chicago, 1962) p. 178.
3. Braddick, H.J.J., The Physics of Experimental Method, (Reinhold Publishing Corp., New York, 1963), 2nd ed., p. 15.
4. Hyde, G.R. and Blount, G.H., "Relative Detector Response to the Irradiance from the Stars," Proc. IRIS, Vol. 8, No. 1, pp 229-237 (January 1963).
5. Uhlenkott, V.H., "Star Brightness and Position Data and Its Application to Star Tracking," Hughes Aircraft Company Rept. SSD3488R.
6. Goldman, S., Frequency Analysis, Modulation and Noise (McGraw-Hill Book Co., Inc., New York, 1948) p. 245.

7. Beall, K. C. and Gee, L. C., editors, "Surveyor I Flight Performance Preliminary Report," Hughes Aircraft Company Rept. SSD681521, pp 167-172 (June 1966).

8. Russell, H. N., Dugan, R. S. and Stewart, J. Q., Astronomy I. The Solar System (Ginn and Co., Boston, 1945), revised ed., p. 358.

9. Van Allen, J. A., "The Geomagnetically Trapped Corpuscular Radiation," Science in Space, edited by Berkner, L. V. and Odishaw, H. (McGraw-Hill Book Co., Inc., New York, 1961), p. 276.

No. 67-587

ORBITAL GYROCOMPASSING HEADING REFERENCE

by

J. L. BOWERS, J. J. RODDEN,
and
E. D. SCOTT
Lockheed Missiles and Space Company
Sunnyvale, California
and
D. B. DeBRA
Stanford University
Stanford, California

AIAA Paper
No. 67-587

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

-- NOTES --

ORBITAL GYROCOMPASSING HEADING REFERENCE

by

J.L. Bowers, Research Engineer
J.J. Rodden, Supervisor
E.D. Scott, Research Specialist

LOCKHEED MISSILES AND SPACE COMPANY
Guidance and Controls, Dept. 62-21
Sunnyvale, California

and

D.B. DeBra, Senior Research Associate and Lecturer
STANFORD UNIVERSITY
Department of Aeronautics and Astronautics
Stanford, California

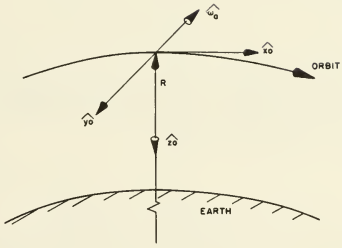
ABSTRACT

Available methods for sensing orientation of a satellite with respect to local orbit coordinates are reviewed, and the importance of gyrocompassing as a heading reference is established. The principles of gyrocompassing are developed by a succession of simple mechanizations. The effect of horizon sensor noise and the suitability of the generated attitude are examined for each system. Shortcomings are used to suggest more sophisticated mechanizations. Some of the simple mechanizations lead to high gas expenditure and high vehicle angular rates when combined with nonlinear controllers. Decoupling loops which are introduced to overcome this problem are shown to be advantageous for instrumentation purposes as well, because they make the gyro gimbal angles directly represent the vehicle attitude. Flight data are presented, verifying the system dynamics. It is concluded that no accuracy is lost due to typical satellite motions by body-mounting the sensors and therefore platform mounting is unnecessary for orbital gyrocompassing.

INTRODUCTION

For earth-pointing satellites, we are principally interested in sensors which give us directly, an indication of orientation error. Sensors that determine orientation with respect to other than the desired local reference, require special calculations and may require the knowledge of the orbit ephemeris as well. Natural references are the vertical, the velocity vector, and the orbital angular velocity vector. Some methods of sensing orientation with respect to these references are shown in Table I. The limitation in orientation knowledge may come from either of two sources: knowledge of the reference itself, or inaccuracies in the sensors.

Information about three axes (yaw, roll, and pitch, see Fig. 1) is needed to provide complete active attitude control. Pitch and roll information is most commonly obtained from a horizon sensor.



ANGLE	AXIS OF ROTATION	BASIS
yaw ψ	$\hat{z}_0$	0, ORBITAL AXES
pitch θ	$\hat{y}_b$	0'
roll ϕ	$\hat{x}_b$	b'
		b, BODY AXES

Note: the $\hat{x}_b$, $\hat{y}_b$, and $\hat{z}_b$ axes are referred to as the roll, pitch, and yaw axes respectively; or simply as the body $\hat{x}$, $\hat{y}$, and $\hat{z}$ axes. The angles are defined as positive in the conventional right hand sense about each axis as tabulated above.

FIG. 1: DEFINITION OF AXES

There are several methods of determining yaw information. Sensing the relative motion of the vehicle with respect to the atmosphere for low-altitude satellites has not been developed and in many cases, would not produce acceptable accuracies. A gyro reference appears attractive, but unless an extremely high performance gyroscope is used, the uncertainties intrinsic in the instrument will limit its usable life. Even though the case of the gyroscope is rotated once each orbital revolution--averaging some drift torques to zero, and the low-g environment reduces some of the drift-producing

TABLE I

ORIENTATION SENSING WITH RESPECT TO "NATURAL" ORBITAL REFERENCES

Reference and Technique	Number of Angles Determined	Reference Knowledge Uncertainty	Sensor Inaccuracy or Noise	Comments
Local Vertical				
Horizon sensor	2	$\approx$ 0.1°	0.1° - 10°	Nature of horizon limits some techniques
Gravity gradient	2	< 1 arc min	< 180°	cf. Diesel 1964, Roberson 1961 Suitable accelerometers may soon be available
Velocity Vector (relative wind)				
Ion gauge	2	$\approx$ 1°	10° ?	May be effected by local gas re-emission and A/C propellants.
Angle of attack meter	1 or 2	$\approx$ 1°	5° - 10° ?	cf. Fry 1966; limited to low altitude.
Coplanar beacon	2	Geometry dependent	< 1°	limited life; complex
Orbital Angular Velocity Vector				
Gyro reference	2	See text	$\psi_g t$ + < 1 arc min	Without correction gyro errors will grow with time.
Gyrocompassing	1	< 1 arc min	0.1° - 10°	The effects of vertical control contribute as well as gyro uncertainties and misalignments. See Table III and text.

torques which limit the gyro's performance on earth--a drift rate on the order of 0.001° per hour would be required to limit the reference uncertainty to 1° after forty days. Furthermore, if the orbital plane is initially misaligned or changes during the satellite lifetime, the gyro-stored angular momentum will change relative to the orbital angular velocity vector, and hence, will not serve as a natural reference for determining the vehicle yaw angle.

This paper discusses orbital gyrocompassing which is very similar to gyrocompassing used in earthbound navigation systems. Gyros can only sense the vehicle angular velocity so it is necessary to have the vehicle controlled to the vertical before a gyro in the vehicle can sense the orbital angular velocity vector. One of the limits in determining the yaw angle is therefore the accuracy with which the vertical is tracked. For the ensuing discussion, the satellite control system uses an horizon sensor as its primary vertical reference. The vehicle may therefore have motions in the roll and pitch axes which are due to the errors in the horizon sensor; these will directly affect the gyrocompassing.

The principles of orbital gyrocompassing will be outlined by considering a sequence of mechanizations starting with the simplest and altering the mechanization as shortcomings appear.

ROLL GYRO ALONE

If a vehicle is constrained to the vertical so that its $\hat{z}$ axis rotates with the orbital angular rate ω_o, then a yaw error will cause a component of the orbital rate to appear along the vehicle roll axis. For small angles, this will be simply $-\omega_o\psi$. Hence, a gyro oriented with its input axis along the roll axis can measure the yaw angle. If the yaw axis control loop is tight, then the output of this gyro will be kept very small and can be considered negligible. The instrument may be a single degree-of-freedom gyro, or an equivalent inertial angular velocity measuring device such as a ring laser gyro or a star field drift sensor. The mechanization is shown schematically in Fig. 2.

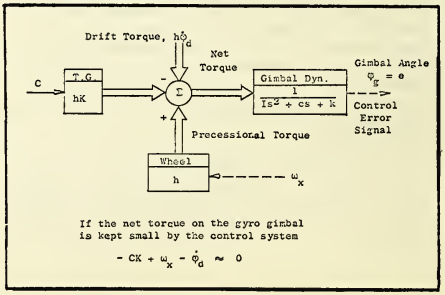


FIG. 2: SINGLE DEGREE-OF-FREEDOM ROLL GYRO FOR YAW CONTROL

For small angles, the angular velocity of an earth-oriented satellite in a circular orbit in terms of the attitude angles is

$$\vec{\omega}_b^b = \begin{bmatrix} \dot{\phi} - \psi\omega_o \\ \dot{\theta} - \omega_o \\ \dot{\psi} + \phi\omega_o \end{bmatrix} \qquad (1)$$

Roll and yaw are coupled due to the orbital motion, but pitch can be considered independently. By assuming tight control, the error signal in Fig. 3 is approximately zero and the roll gyro equation is reduced to

$$\dot{\phi} - \psi\omega_o = \dot{\Phi}_d \quad . \qquad (2)$$

In this way, we can solve for the yaw angle in terms of the forced motion introduced by the control and disturbance torques about the roll axis and the instrument uncertainty torque that produces the drift, $\dot{\Phi}_d$

$$\psi = (\dot{\phi} - \dot{\Phi}_d)/\omega_o . \qquad (3)$$

This type of sensing would be totally unacceptable. Assuming the roll angle varies sinusoidally with a frequency of 10^{-1} rad/sec, and is never larger than 10^{-2} rad, the yaw attitude would be forced to oscillate with an amplitude of about 1 rad at the same frequency.

Decoupling

Since the trouble comes from the roll motion, one can attempt to correct for this error by cancelling the roll signal from the gyro output. Using a rate-integrating gyro, a signal from a roll horizon sensor is subtracted from the gyro output as described in Eqs. 4. The gyro time constant has been neglected ($I\ddot{\Phi}_g \ll c\dot{\Phi}_g$).

$$c\dot{\Phi}_g/K = h(\dot{\phi} - \omega_o\psi) - h\dot{\Phi}_d \qquad (4a)$$

$$e = \Phi_g - (\phi + u_{HS})$$

Setting $K = c/h$ and assuming tight control, the yaw attitude is

$$\psi = (-\dot{u}_{HS} - \dot{\Phi}_d)/\omega_o \quad . \qquad (4b)$$

This does not solve the problem, however, as the horizon sensor noise $\dot{u}_{HS}$ now effects the error signal e as $\dot{\Phi}$ had before. Furthermore, it now makes the control loop for the vertical axis contain three integrations (one in the gyro, and two for the vehicle). When compensated, this would be a conditionally stable control system.

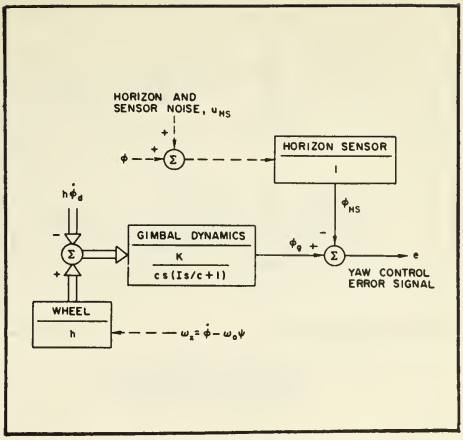


FIG. 3: SINGLE DEGREE-OF-FREEDOM ROLL GYRO WITH ROLL DECOUPLING FOR YAW CONTROL.

THE NEED FOR TIGHTER YAW CONTROL

The yaw information measured by the roll gyro is quite small. Therefore, when a suitable control gain is supplied for yaw control, the other signals, which are large compared to the yaw information, introduce a large error. It would therefore be desirable to have a tight yaw control loop with a reasonably long memory which could be gently steered by the gyrocompass information without responding to the noise. This can be accomplished using a second gyroscope as the memory for the tight yaw control in addition to the roll gyroscope which obtains the gyrocompass information.

When two gyros are available, it is frequently desirable to use them both for tight control of their respective axes, letting the horizon sensor correct drifts in the roll axis gyro instead of acting as the basic sensor. When this is done, the output of the roll gyro--which is the error signal for roll control--is kept negligibly small. The gyrocompass information is no longer available in the roll gyro output, but information sensed by the roll gyro influences the roll attitude. Therefore, the horizon sensor signal contains the yaw attitude information sensed by the roll gyro and can be used to erect the yaw gyro as shown in the mechanization of Fig. 4. The gyro equations for two instruments can then be written as shown in Eqs. 5 and 6.

Roll gyro: $h(\dot{\phi}-\omega_o\psi) - h\dot{\Phi}_d + K_\phi h(\phi+u_{HS}) = c\dot{\Phi}_g$; (5)

Yaw gyro: $h(\dot{\psi}+\omega_o\phi) - h\dot{\Psi}_d + K_\psi h(\phi+u_{HS}) = c\dot{\Psi}_g$. (6)

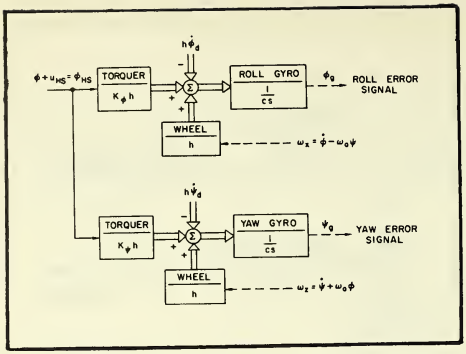


FIG. 4: GYROCOMPASS WITH TWO SINGLE-DEGREE-OF-FREEDOM RATE INTEGRATING GYROS.

Taking the Laplace transform and solving for the vehicle attitude angles gives

$$\begin{bmatrix} \Phi \\ \Psi \end{bmatrix} = \frac{1}{\Delta(s)} \begin{bmatrix} s & \omega_o \\ -(\omega_o + K_\psi) & (s + K_\varphi) \end{bmatrix} \begin{bmatrix} \varphi(0) + \frac{\dot{\Phi}_d}{} - K_\varphi U_{HS} + \frac{c}{h}\dot{\Phi}_g \\ \psi(0) + \dot{\Psi}_d - K_\psi U_{HS} + \frac{c}{h}\dot{\Psi}_g \end{bmatrix} \quad (7)$$

The natural behavior implied by the characteristic equation

$$\Delta(s) = s^2 + K_\varphi s + \omega_o(\omega_o + K_\psi) = 0, \quad (8)$$

is a fairly slow response with a typical settling time constant of at least 10 sec. It represents the dynamics of the gyrocompassing and does not include any of the vehicle dynamics associated with the tight control loops in the roll and yaw axes. Figure 5 shows a plot of the characteristic roots as a function of the two gain parameters over the usual range of interest. A typical gain setting might be 10^{-2} rads/sec for K_φ and 3×10^{-2} rads/sec for K_ψ. The gyrocompass erection dynamics would then be critically damped and would have a settling time of several minutes. Final value errors are given in Table II.

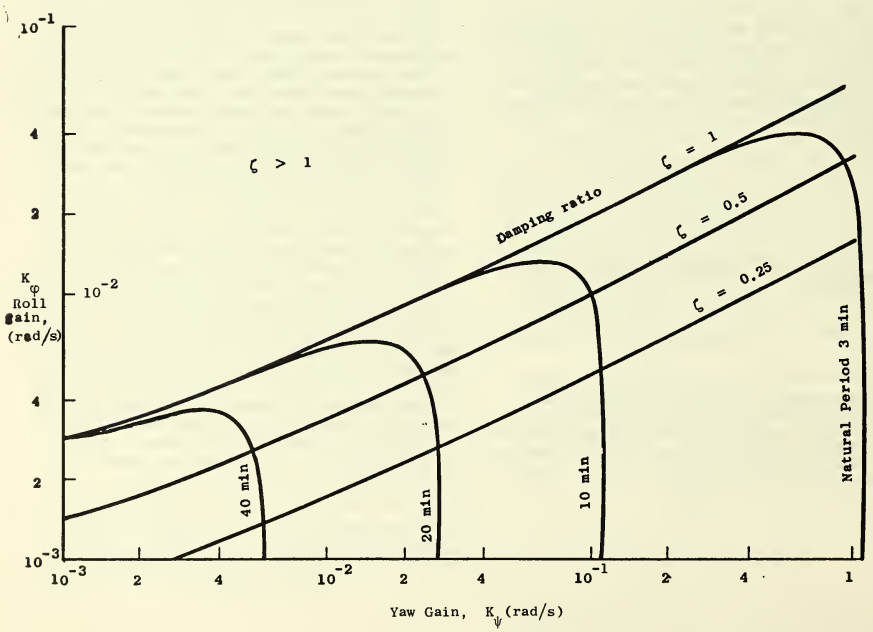


FIG. 5: NATURAL PERIOD AND DAMPING OF GYROCOMPASS AS A FUNCTION OF GAINS IN FIG. 4.

TABLE II

FINAL-VALUE GYROCOMPASSING ERRORS

Assuming Constant Values for $\dot{\Phi}_d$, $\dot{\Psi}_d$, u_{HS}, and the Error Signals Φ_g, Ψ_g, or α_Φ, α_Ψ :

Error Source	Contribution To $\Phi(\infty)$	Contribution to $\Psi(\infty)$	Comment
Two Gyros, see Fig. 3.			
Roll gyro drift	0	$-\dot{\Phi}_d/\omega_o$	Basic gyrocompass error
Yaw gyro drift	$\dot{\Psi}_d/(\omega_o + K_\Psi)$	$\dot{\Psi}_d K_\Phi/\omega_o(\omega_o + K_\Psi)$	
Roll horizon sensor error, misalignment and horizon noise	$-u_{HS}K_\Psi/(\omega_o + K_\Psi)$	$u_{HS}K_\Phi/(\omega_o + K_\Psi)$	
Roll gyro misalignment about output axis	0	$-\Phi_e$	
Yaw gyro misalignment about outputs axis	$\Psi_e K_\Psi/(\omega_o + K_\Psi)$	$\Psi_e K_\Phi/(\omega_o + K_\Psi)$	
Roll or yaw gyro-angle offset due to external torque and deadbands	0	0	Gyro integration removes the error.
Two DOF Gyro, see Fig. 5.			
Same except:			
Roll gyro-angle offset due to external torque and deadband	$\alpha_\Phi\omega_o/(\omega_o + K_\Psi)$	$\alpha_\Phi K_\Phi/(\omega_o + K_\Psi)$	
Yaw gyro-angle offset due to external torque and deadband	0	α_Ψ	
Skewed Gyro, see Fig. 6.			
Gyro drift		$\dot{\Phi}_d/\omega_o \cos\beta$	Assuming perfect decoupling gains.
Horizon Sensor error, misalignment and horizon noise.		$u_{HS}K/\omega_o \cos\beta$	

As in the case of the single axis gyro, the assumption has been made that one is interested in the vehicle attitude when the control is tight and hence, the error signals (the gyro gimbal angles in this case) are negligible. Later we will discuss the importance of the error signals representing the vehicle attitude and it will be more interesting to solve Eq. 10 for the error signals Φ_g and Ψ_g

Although two single-degree-of-freedom rate-integrating gyros were used to derive these results, a two-degree-of-freedom gyro will serve just as well. Gimbal angles for a free rotor gyro are defined in the sketch of Fig. 6. Equations 9 and 10 can be compared with Eqs. 6 and 7. In Eqs. 9 and 10, the

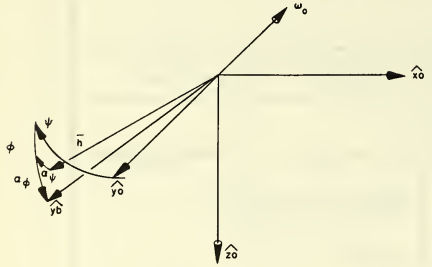


FIG. 6: TWO-DEGREE-OF-FREEDOM GYRO GIMBAL ANGLES

torque equation about yaw is written first. It produces a roll precession and is labeled the "roll equation." Roll torques on the gyro produce a negative yaw precession and hence, the second is the yaw equation. The gyro response determines the name (and symbols) used for each instrument channel rather than the axis of the torque which produces it. Expanding $\dot{\bar{h}} = T$

$$\text{roll: } (\dot{\phi} - \dot{\alpha}_\phi)h - \omega_o(\psi - \alpha_\psi)h = -K_\phi h(\phi + u_{HS}) + h\dot{\phi}_d \tag{9}$$

$$\text{yaw: } -(\dot{\psi} - \dot{\alpha}_\psi)h - \omega_o(\phi - \alpha_\phi)h = K_\psi h(\phi + u_{HS}) - h\dot{\psi}_d$$

$$\begin{bmatrix} \phi \\ \psi \end{bmatrix} = \frac{1}{\Delta(s)} \begin{bmatrix} s & \omega_o \\ -(\omega_o + K_\psi) & (s + K_\phi) \end{bmatrix} \tag{10}$$

$$\times \begin{bmatrix} (\phi - \alpha_\phi)(0) + \dot{\phi}_d - K_\phi U_{HS} + sA_\phi - \omega_o A_\psi \\ (\psi - \alpha_\psi)(0) + \dot{\psi}_d - K_\psi U_{HS} + \omega_o A_\phi + sA_\psi \end{bmatrix}.$$

In a free rotor two-degree-of-freedom gyro (or one with gimbals that introduce negligible restraint) the initial conditions are with respect to the orbit normal as would be expected. The terms involving the vehicle angles with respect to the gyro, α_1 and α_2, are used as the control error signal.

The output of the two rate-integrating gyros approximately equal the change in vehicle orientation. The responses are fast and quiet so they can be used for tight control. In the single gyro case, where the input axis was along the roll axis, no short-term yaw information was obtained. Consider the possibility of changing the orientation of the input axis so that it senses some of the angular velocity about the vehicle yaw axis as well as about the roll axis. The output of a rate-integrating gyro in the orientation defined in Fig. 7 will then contain both yaw and roll information as well as the integral of these angles. Again, let us

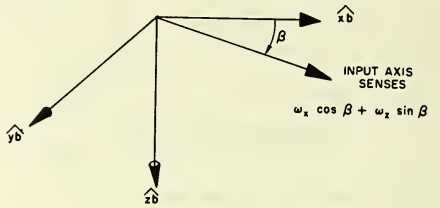


FIG. 7: ORIENTATION OF GYRO INPUT AXIS IN THE ROLL-YAW PLANE FOR GYROCOMPASSING WITH ONE SINGLE DEGREE-OF-FREEDOM RATE-INTEGRATING GYRO.

try and remove the roll information by obtaining a signal from the horizon sensor as we did before. Now the yaw control will not be as difficult as before because the error signal will contain the yaw angle as well as its integral. The mechanization of Fig. 8 permits one to do away with some of the principal drawbacks of a single roll-axis-gyro mechanization, although it does not permit the tight control in roll and yaw that can be obtained with two gyros. The equations for this system are

$$c\dot{\sigma} = h(\omega_x \cos\beta + \omega_z \sin\beta + K_1\phi + K_1 u_{HS} - \omega_d - K_3\sigma) \tag{11}$$

$$e = \sigma + K_2(\phi + u_{HS}) \tag{12}$$

$$e_\psi = K_4 e \tag{13}$$

The gain loops K_1 and K_2 remove roll motion contamination from the yaw control signal. First assume $K_3 = 0$, then assuming tight control so that the error signal is negligible and solving for the yaw angle, one obtains

$$\psi = -\left\{\Phi[s(\cos\beta + K_2 c/h) + \omega_o \sin\beta + K_1] + U_{HS}[sK_2 c/h + K_1] - \omega_d\right\}/(s\sin\beta - \omega_o\cos\beta). \tag{14}$$

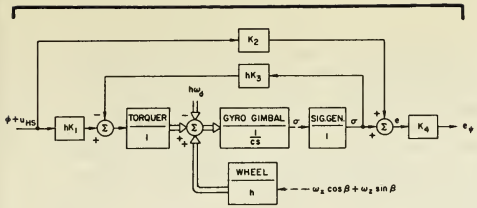


FIG. 8: SKEWED/CANTED SINGLE GYRO HEADING SENSOR WITH DECOUPLING.

Now, by adjusting the two gain parameters K_1 and K_2 to satisfy

$$K_1 = -\omega_o \sin \beta \,, \qquad (15)$$

$$K_2 = -h \cos \beta / c \,, \qquad (16)$$

the coefficients of Φ can be made zero. In practice, they might be reduced by, say, 10^{-2}, and the forced errors due to roll motion can be made a tolerable size. To make the gyrocompass settle rapidly, one desires a small angle β, and cot β must be negative for stability. The response to horizon sensor noise u_{HS} becomes excessive for these gains at high frequencies when β is small, and at low frequencies for values of β approaching $-90°$. Furthermore, near $-90°$, the gyro cannot sense as much orbital rate for a given yaw angle and hence, the error due to gyro drift rate becomes large. These two effects are shown in Fig. 9 where a noise sine wave with an amplitude of 10^{-2} rad has been assumed. The dashed curves show the effect of noise whose amplitude is dependent upon frequency, f (note $n \equiv \omega_o$ on this figure). A decreasing amplitude above a frequency of 100 ω_o was assumed. As an example, a gyro with a drift rate $\delta = 10^{-2}\omega_o$ and an horizon sensor with noise of 10^{-2} rad amplitude could perform yaw control with an accuracy of better than 1° if the angle β were chosen at approximately $-40°$ (see also Table II).

The foregoing discussion was intended to give some background on the principles of gyrocompassing. The key points are:

- The gyrocompassing information is a small signal compared with typical roll rates which contaminate it;
- The gyrocompassing signal is relatively free of noise only for bandwidths that are too small to be of interest for direct yaw control;
- A rate integrating, or a displacement gyro, can supply attitude information with sufficient bandwidths for tight attitude control but has only a short-term memory due to its drift;
- Mechanizations are possible in which the correction of drift in gyros used for short-term memory can be obtained using a horizon sensor for pitch and roll information and gyrocompassing for yaw information;
- It is possible to combine the function of short-term memory and gyrocompassing in a single gyro by canting its input axis between the roll and yaw axes.

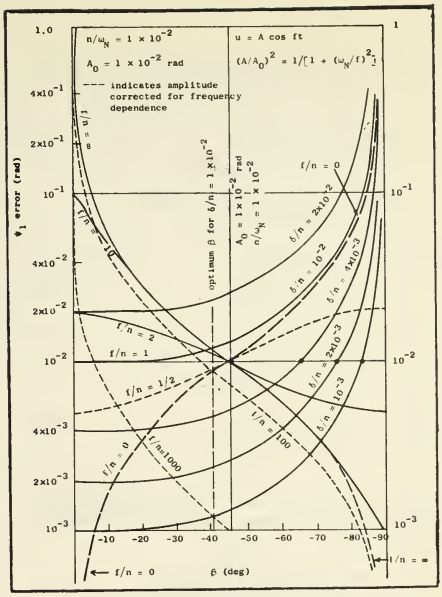


FIG. 9: THE EFFECT OF HORIZON SENSOR NOISE AND GYRO DRIFT ON SINGLE-GYRO GYROCOMPASSING.

These ideas underlie all gyrocompassing. There is an additional need which will be overriding in the ensuing discussion. The output of the gyros in the above mechanizations are only approximately the attitude error. It is desirable both for control and instrumentation to have the gyro outputs accurately represent the vehicle attitude. This can be done by adding "decoupling loops." The motivation and derivation for these decoupling loops will be given, then a discussion of the control of a vehicle using a horizon sensor and gyrocompassing will illustrate their importance.

DECOUPLING THE SKEWED GYRO

Setting K_1 and K_2 as prescribed in Eqs. 15 and 16, the roll errors can be removed. However, the yaw control signal is not equal to the yaw Euler angle. The signal contains a component of the integral of yaw. With one additional decoupling loop, the dynamics in the transfer function from ψ to e_ψ may be removed, making the control output a measure of the true yaw Euler angle. This loop is the gain K_3 shown in Fig. 8. Being interested in the error signal, Eqs. 11 to 13 are solved for e_ψ giving

67-587

$$E_\psi = K_4 \left\{ \Phi \frac{s(\cos\beta + K_2 \frac{c}{h}) + (\omega_o \sin\beta + K_1 + K_2K_3)}{sc/h + K_3} + \Phi \frac{s\sin\beta - \omega_o \cos\beta}{sc/h + K_3} + \frac{U_{HS}(sK_2\frac{c}{h} + K_1 + K_2K_3) - \omega_d}{sc/h + K_3} \right\} \quad (17)$$

Then e_ψ will follow ψ for

$$K_1 = -\omega_o/\sin\beta \quad (18)$$

$$K_2 = -h\cos\beta/c \quad (19)$$

$$K_3 = -\omega_o c/h \tan\beta \quad (20)$$

and finally e_ψ is properly scaled to give ψ when

$$K_4 = -c/h \sin\beta \, . \quad (21)$$

With both the roll contamination and the dynamics associated with yaw sensing removed, the gyro output is given by

$$E_\psi = \Psi + \frac{-s\cot\beta + \omega_o}{s + \omega_o \cot\beta} U_{HS} - \frac{\omega_d}{\sin\beta(s+\omega_o\cot\beta)} \quad (22)$$

Note that with the addition of the K_3 decoupling loop, the gain K_1 is different from the value required without the K_3 loop (compare Eq. 18 with Eq. 15).

DECOUPLING LOOPS FOR A TWO-GYRO MECHANIZATION

The two-gyro system previously developed (Fig. 3) does not indicate the true roll and yaw Euler angles. The kinematic cross-coupling of the vehicle roll and yaw motions (Eq. 1) is reflected in contamination of the gyro outputs so that each gyro contains components of both roll and yaw Euler angles. The use of the roll horizon sensor for gyrocompassing and roll gyro drift correction further distorts and contaminates the gyro outputs. In a satellite like the Agena, shown in Fig. 10, this has a significant effect on the attitude control performance. The use of "decoupling loops" provides a means for circumventing these difficulties, and enables the two gyros to measure directly the Euler angles. These are electrical loops interconnecting the two gyros in such a way as to reconstruct the Euler angles as gyro outputs. The gyro and horizon sensor system are shown in Fig. 11, with decoupling loop gains indicated by K_1, K_2, and K_3 .

The gyros sense the angular rate about the axis normal to their spin and output axes. When the gimbal angle is not zero this does not coincide with the sensitive axis indicated by the gyro case. When the nominal direction of the spin axis is normal to the orbit plane as in the case of the yaw and roll gyros of the Agena, each additional contribution of the orbital rate is sensed proportional to the gyro gimbal angle. This "loose caging" effect is neglected in the following discussion.

Decoupling is of two types: kinematic decoupling and horizon sensor decoupling. The kinematic decoupling consists of the two loops indicated by K_1 and K_2 in Fig. 12. These loops connect the gyros to form a model of the vehicle kinematics. Thus, given the body rates ω_x and ω_z as inputs to the model, K_1 and K_2 may be set so that the model behaves in identical fashion to the vehicle kinematics and produces as outputs, the roll and yaw Euler angles φ and ψ (ignoring gyro drift and higher order terms).

Due to gyro drifts, however, the model is not perfect and a means of drift correction must be provided. This is done by gyrocompassing as previously discussed, where a correction signal for the roll and yaw gyros is obtained from the roll horizon sensor signal. However, simply connecting the roll horizon sensor signal to the roll and yaw gyros introduces cross-coupling so that the gyros no longer indicate the roll and yaw Euler angles. It is necessary, therefore, to generate a gyrocompassing correction signal which is a function only of the gyro drift terms, not the vehicle Euler angles.

Such a signal may be derived by taking the difference between the roll gyro and the roll horizon sensor and using this difference as a drift trim signal for the gyros. The loop equations for the interconnected gyros alone shown in Fig. 12 are

$$\begin{bmatrix} \Phi_g(s) \\ \Psi_g(s) \end{bmatrix} = \left\{ \begin{bmatrix} (s^2 + \omega_2\omega_o) & (K_2 - \omega_o)s \\ (\omega_o - K_1)s & (s^2 + K_1\omega_o) \end{bmatrix} \begin{bmatrix} \Phi(s) \\ \Psi(s) \end{bmatrix} + \begin{bmatrix} s\omega_{x_d}(s) + K_2\omega_{z_d}(s) \\ -K_1\omega_{x_d}(s) + s\omega_{(s)} \end{bmatrix} \right\} \cdot \frac{1}{s^2 + \omega_o^2} \quad (23)$$

By taking $K_1 = K_2 = \omega_o$, the gyro signals may be "decoupled," eliminating the kinematic cross-coupling, yielding

$$\Phi_g(s) = \Phi(s) + \frac{s\omega_{x_d}(s) + \omega_o\omega_{z_d}(s)}{s^2 + \omega_o^2} \quad (24)$$

$$\Psi_g(s) = \Psi(s) + \frac{-\omega_o\omega_{x_d}(s) + s\omega_{z_d}(s)}{s^2 + \omega_o^2} \quad (25)$$

From Eq. 24, it is seen that by differencing the roll horizon sensor signal and the roll gyro signal, one obtains a net signal which is simply a function of the gyro drifts (uncontaminated by components of the roll and yaw Euler angles). This signal differencing is equivalent to taking $K_3 = 1$ in Fig. 11. This difference is introduced at the roll and yaw gyro torquers (with respective gains of K_φ and K_ψ) to provide a drift correction to the gyros.

With these connections made, and with $K_1 = K_2 = \omega_o$, the gyro outputs--in terms of the Euler angles and the drift rates--become

$$\begin{bmatrix} \Phi_g(s) \\ \Psi_g(s) \end{bmatrix} = \frac{\begin{bmatrix} s^2 + K_\varphi s + \omega_o(K_\psi + \omega_o) & 0 \\ K_\psi(1 - K_3)s - (K_3 - 1)\omega_o K_\varphi & s^2 + K_3 K_\varphi s + \omega_o(K_3 K_\psi + \omega_o) \end{bmatrix} \begin{bmatrix} \Phi(s) \\ \Psi(s) \end{bmatrix}}{s^2 + K_3 K_\varphi s + \omega_o(K_3 K_\psi + \omega_o)}$$

$$+ \frac{\begin{bmatrix} s\omega_{x_d}(s) + \omega_o \omega_{z_d}(s) \\ -(K_3 K_\psi + \omega_o)\omega_{x_d}(s) + (s + K_3 K_\varphi)\omega_{z_d}(s) \end{bmatrix}}{s^2 + K_3 K_\varphi s + \omega_o(K_3 K_\psi + \omega_o)} \qquad (26)$$

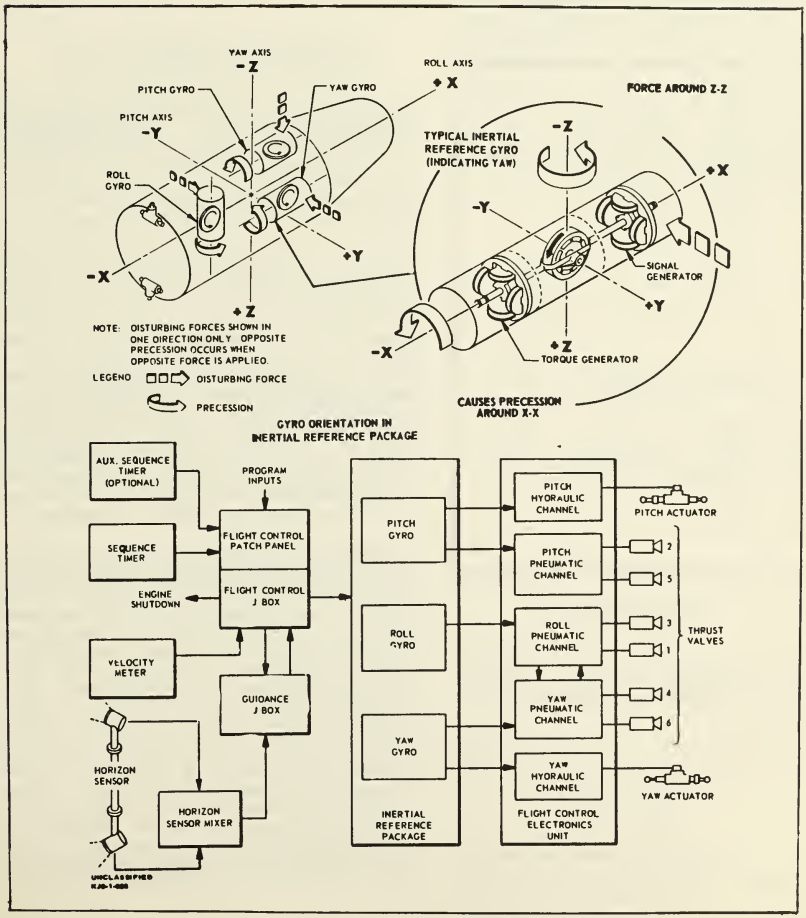


FIG. 10: AGENA GUIDANCE AND FLIGHT CONTROL SYSTEM

67-587

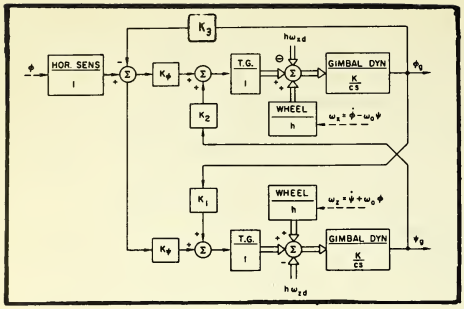


FIG. 11: HORIZON SENSOR DECOUPLING

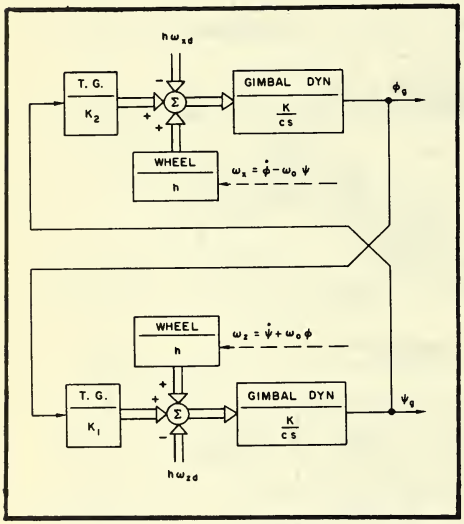


FIG. 12: KINEMATIC DECOUPLING

Taking $K_3 = 1$, the off-diagonal term in the first matrix is zero and therefore the the cross-coupling is removed and the diagonal terms cancel with the denomination giving

$$\Phi_g(s) = \Phi(s) + \frac{s\omega_{x_d}(s) + \omega_o\omega_{z_d}(s)}{s^2 + K_\varphi s + \omega_o(K_\psi + \omega_o)} \qquad (27)$$

$$\Psi_g(s) = \Psi(s) + \frac{-(K_\psi+\omega_o)\omega_{x_d}(s)+(s+K_\varphi)\omega_{z_d}(s)}{s^2 + K_\varphi s + \omega_o(K_\psi + \omega_o)} \qquad (28)$$

It will be noticed that the characteristic polynomial is the same as that obtained in Eq. 8 for gyrocompassing without decoupling. The decoupling loops alter the zeros associated with the transfer function of the observer, so as to permit direct observation of the desired variables φ and ψ .

After the decay of the transients in the error terms, assuming constant gyro drifts $\omega_{x_d}(s) = \omega_{x_d}/s$ and $\omega_{z_d}(s) = \omega_{z_d}/s$, the gyro outputs become the gyro errors of Table II plus the vehicle attitude

$$\varphi_g(\infty) = \varphi(\infty) + \frac{1}{K_\psi + \omega_o}\omega_{z_d} \qquad (29)$$

$$\psi_g(\infty) = \psi(\infty) - \frac{1}{\omega_o}\omega_{x_d} + \frac{K_\varphi}{\omega_o(K_\psi + \omega_o)}\omega_{z_d} \qquad (30)$$

THE USE OF ERROR SIGNALS FOR CONTROL

In a satellite control system, the error signal (gyro angles) are not zero. Frequently they are permitted to wander within a deadband and control action is taken in discrete pulses only when the error signal exceeds the deadband. In this case, both stability and performance of the satellite attitude control loop are enhanced by the availability of φ and ψ as control error signals. A system with deadband and pulsed control permits the existence of a steady state limit cycle in the roll-yaw attitude motion. Without decoupling loops, the maximum amplitude of the yaw limit cycle motion may be significantly larger (50% or more) than the control deadband, because the yaw gyro output is reduced by roll horizon sensor torquing. In addition, the roll and yaw body rates in the limit cycle are substantially higher (by orders of magnitude in some cases) than the theoretical minimum associated with a single pulse or minimum impulse limit cycle, resulting in a greater expenditure of propellant. This form of performance degradation results from the gyro angles not adequately representing the vehicle orientation. Daughton [1966] has explained some anomalous behavior on the Agena that resulted from this coupling phenomenon.

Figure 13 shows two root loci of the gyrocompassing loop with yaw autopilot gain as the parameter, the first without and the second with decoupling loops. A linear system is assumed. The roll dynamics produces a pole at the origin and another pole on the real axis at a frequency proportional to the roll horizon sensor gain. The vehicle produces a double pole at the origin. The gyrocompassing appears as low frequency zeros. The coupling of the low frequency dynamics (resulting in the three poles at the origin) cause the roots emanating from the origin to pass into the right half of the s-plane for low autopilot gains. Since one effect of introducing a dead zone is to lower the effective system gain, the system is forced to limit cycle at approximately the frequency where the locus crosses the jω axis for low gains. When the frequency of a limit cycle is fixed in this way, gas expenditure increases with deadband and with vehicle inertia! With decoupling, the poles

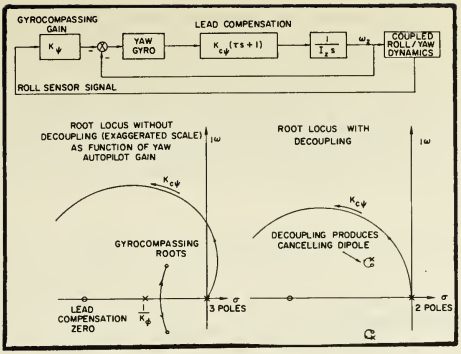


FIG. 13: THE EFFECT OF DECOUPLING ON CONDITIONAL STABILITY.

due to the roll dynamics are shifted to cancel the gyrocompassing roots. With this cancellation, the yaw dynamics are that of a simple inertia with lead compensation, and the root locus remains entirely in the left-half plane, insuring stability for all gains.

The incorporation of the three decoupling loops described above enables the gyros to measure directly roll and yaw attitude. Control may now be exercised directly over the two variables of interest; hence, maximum roll and yaw excursions are constrained to the deadband values, and since the conditional stability is removed, the controller fires only a single minimum-impulse bit when the deadband is exceeded. Body rates in the limit cycle are thus simultaneously reduced to their minimum values. However, the decoupling loops permit vehicle offsets in the presence of steady external torques unless an additional integration is included in the controller.

Table III summarizes simplified relations of the gyro signals in terms of the Euler angles for various combinations of decoupling loops. With no decoupling, the roll gyro signal contains a component proportional to the integral of the roll angle, and the yaw gyro similarly contains the integral of yaw. When the gyro signals actuate the control, a steady external torque in either axis requires a corresponding steady gyro output. Due to the existence of the integral control without the decoupling loops, the Euler angles are driven to zero in the presence of disturbance torques. With decoupling,

TABLE III

GYRO-SIGNALS--ATTITUDE RELATIONS FOR SEVERAL DECOUPLING CONFIGURATIONS

Decoupling Loops	Roll Gyro Signal $\Phi_g(s)$	Yaw Gyro Signal $\Psi_g(s)$
None	$\left(1 + \frac{K_\Phi}{s}\right)\Phi - \frac{\omega_o}{s}\Psi$	$\left(1 + \frac{\omega_o^2}{s}\right)\Psi + \frac{K_\Psi + \omega_o}{s}\Phi$
K_2 only	$\left[1 + \frac{K_\Phi}{s} + \frac{\omega_o(K_\Psi + \omega_o)}{s^2}\right]\Phi$	$\Psi + \left[\frac{K_\Psi + \omega_o}{s}\right]\Phi$
K_1 only	$\left(1 + \frac{K_\Phi}{s}\right)\Phi - \frac{\omega_o}{s}\Psi$	$\left(1 + \frac{\omega_o^2}{s^2}\right)\Psi + \left(\frac{K_\Psi}{s} - \frac{\omega_o K_\Phi}{s^2}\right)\Phi$
K_1 and K_2	$\frac{s^2 + K_\Phi s + \omega_o(K_\Psi + \omega_o)}{s^2 + \omega_o^2}\Phi$	$\Psi + \left(\frac{K_\Psi s - \omega_o K_\Phi}{s^2 + \omega_o^2}\right)\Phi$
All	Φ	Ψ

however, the gyro outputs are equal to the Euler angles. A constant gyro output, therefore, requires a steady attitude offset under an external torque.

FLIGHT RESULTS

Figure 14 shows flight performance of an early gyrocompassed Agena. The vehicle is reorienting itself to the orbit reference frame from an initial attitude offset. The gyro configuration is partially decoupled in roll and yaw. The gain loop from the yaw gyro to the roll torquer was not included. The independence of the pitch motion from roll and yaw is demonstrated. The decoupled roll channel shows the horizon sensor and gyro signals for a sensor gain of 0.0167 sec^{-1} in yaw. The difference between the roll gyro and sensor signal is the yaw torquing rate which shows time constants as

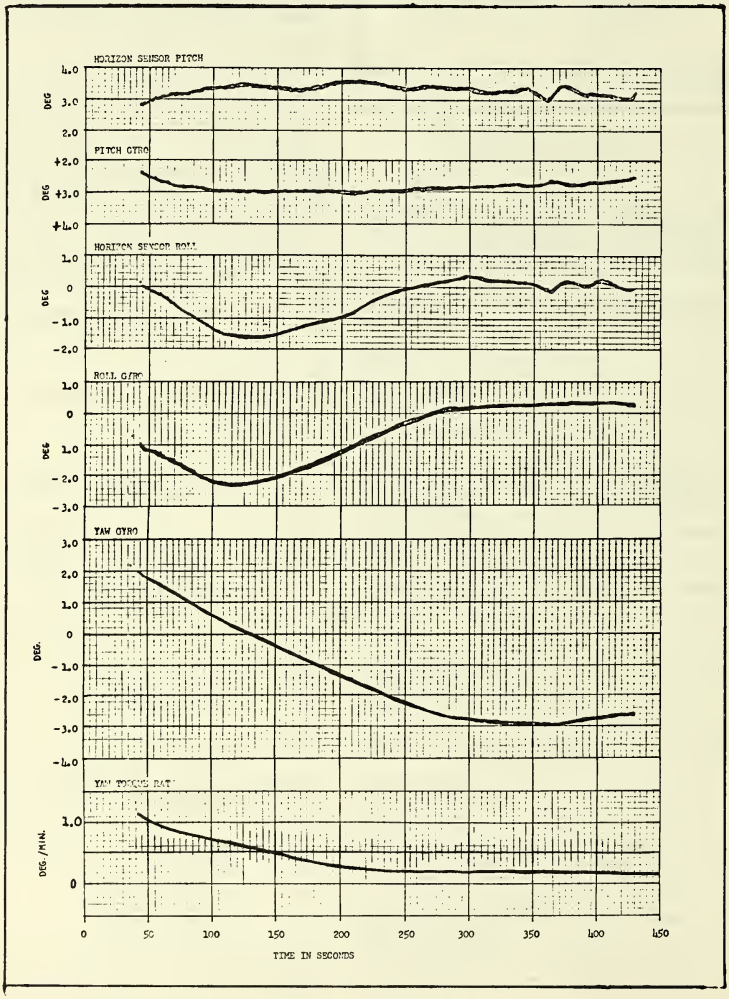


FIG. 14: GYROCOMPASS FLIGHT DATA

approximately 75 sec and a much much longer value. These are within acceptable tolerance with the predicted time constants of 65 and 900 sec respectively. It should be noted that the gyro signals do not reflect heading reference errors because they are the attitude control error signals and are kept within the deadband (in this case ± 3°). The heading reference error information forced a roll attitude which can therefore be obtained as the difference between the roll gyro and horizon sensor signals.

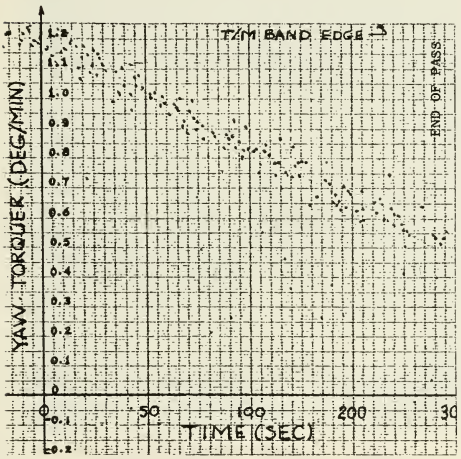


FIG. 15: AGENA TYPICAL GYROCOMPASSING TRANSIENT RESPONSE.

In Fig. 15, the data show another vehicle recovering from a transient after which the yaw heading reference must be re-established. The gyro gimbal angles remain within the deadband after the vehicle dynamics have settled as they are the error signals for control. The gyrocompass erection signal appears as the yaw gyro torque rate. The gyrocompassing gains on this flight were $K_\varphi = 1/180\ sec^{-1}$ and $K_\psi = 1/90\ sec^{-1}$ with the same partial decoupling as the vehicle discussed above. The expected erection time constant of 6 minutes correlates well with the observed approximately 330 sec time constant of the data.

DISCUSSION

The several mechanizations discussed have shown that a heading reference can be achieved by angular velocity measurements in several ways. The utility of the error signals for control is improved by adding decoupling loops which permit sensing the Euler angles directly. In some cases, these loops result in a considerable saving in propellant required for limit cycling by removing the conditional stability at low gains.

Orbital gyrocompassing is very similar in concept and behavior to the earth-navigator gyrocompasses described by Cannon [1961]. The second order system he discusses has exactly the same characteristic equation as the two-gyro, orbital gyrocompassing system. While the horizon sensor signal could be more heavily filtered to improve the gyrocompassing (and hence be more analogous to the third order system discussed by Cannon [1961], the filtering is typically kept at about 1 rad/sec to prevent introducing too much phase shift which is destabilizing to basic "tight" attitude control loop.

The systems discussed here have all involved instruments which are mounted in the vehicle body. A more sophisticated approach is to isolate the instruments on a platform but still introduce the same horizon sensor signal for erection of the roll and yaw gyros. Gordon [1964] describes a platform system and discusses two mechanizations giving the principal error terms. Because of the low-g environment and small angular accelerations, body-mounted gyrocompassing systems exhibit no additional errors to those listed by Gordon. The equivalent performance and need for high reliability make the body-mounted systems attractive. There does not appear to be any significant advantage for use of a platform in orbital gyrocompassing applications as there is for conventional navigation systems.

ACKNOWLEDGEMENTS

The authors wish to thank B. Procopio and C. Hammes for their assistance in securing examples of flight data and to acknowledge the assistance of A. J. Daughton in the decoupling discussion on forced limit cycles.

REFERENCES

Cannon, R.H. Jr., "Alignment of Inertial Guidance Systems by Gyrocompassing—Linear Theory," J. of the Aerospace Sci., vol. 28, No. 11, Nov 1961

Daughton, A.J., "Design Criteria for the Agena Flight Control System," LMSC Internal Memo No. A813278, Lockheed Missiles and Space Corp., Sunnyvale, Calif., 20 Dec 1966, (unclassified).

Diesel, J.W., "A New Approach to Gravitational Gradient Determination of the Vertical," AIAA J. vol. 2, pp. 1189-1196, 1964.

Frye, William E., "Analysis of a Satellite Angle-of-Attack Sensor," 1st Aerospace Mechanisms Symposium, May 19-20, 1966, University of Santa Clara, Santa Clara, Calif.

Gordon, R.L., "An Orbital Gyrocompass Heading Reference for Satellite Vehicles," 1st AIAA Annual Meeting, Jun 29-Jul 2, 1964, Washington, D.C., AIAA Paper No. 64-238.

Roberson, R.E., "Gravity Gradient Determination of the Vertical," ARS J., vol. 31, pp. 1509-1515, 1961.

-- NOTES --

No. 67-588

APOLLO GUIDANCE, NAVIGATION AND CONTROL SYSTEM GYRO RELIABILITY

by

JOHN E. MILLER and JULIUS FELDMAN
Massachusetts Institute of Technology
Cambridge, Massachusetts

AIAA Paper
No. 67-588

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

APOLLO GUIDANCE, NAVIGATION AND CONTROL SYSTEM GYRO RELIABILITY

John E. Miller, Associate Director
Julius Feldman, Group Leader

Instrumentation Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

ABSTRACT

The reliability of the Apollo Guidance, Navigation and Control (GN&C) system depends to a large measure upon the reliability of the inertial reference gyro. This fact was recognized at the inception of the Apollo GN&C system design. A proven design, the Polaris Mod II Inertial Reference Integrating Gyroscope (IRIG), was used as the starting point. Although the signal and torque microsyns were redesigned to meet the frequency and computer interface requirements, the critical reliability components, the float and wheel package, were unchanged from a design already in production.

This paper presents the method used to increase the operational reliability of the Apollo gyro by a factor of twenty despite the fact that the wheel assembly reliability had exploited the state of the art. The results of the reliability program led to a method of forecasting the occurence of the primary failure mode of the gyroscope, the failure of the wheel bearing assembly. This failure prediction method, called the Delta 25 test has been heuristically derived. Application of this test to a sample of gyros confirmed that it was reasonably accurate in its ability to predict a failure with sufficient operating time with good performance to complete a lunar mission.

Finally, a second method of failure prediction has been derived which has the same usefulness as the Delta 25 test but with increased efficiency. The method, called the F criterion, is the RMSSD (root mean squared value of successive differences) and is applied to the last N data points. A comparison of the F method with the Delta 25 test has been made for a sample of the Apollo gyros.

The overall gyro reliability for the Apollo GN&C gyro is calculated. The history of system performance is also shown and some of the more interesting predicted failure examples are shown in detail.

Introduction

The mission success of Apollo is dependent upon the reliability of the Guidance Navigation and Control System. Recognized at design inception was the dependence of GN&C reliability upon the inertial reference gyroscope reliability. Insufficient time existed to create a special or new gyroscope design. It was necessary to select a design with good production maturity and with adequate performance capability to meet the mission requirements. The objective was to utilize gyros on each flight which had a mean time before failure of 100,000 hours or greater. This failure rate was to be experienced for a set of three instruments for time period of wheel operation of about 300 hours — a time representative of the final checkout period and the flight. Thus it was necessary to be able to predict with good accuracy the reliability of the instrument for a limited time period. The problem then is one of measuring the performance and the symptoms of performance over the life of a quantity of instruments and using this information to predict performance for the next time period and thus predict reliability. The prediction was slightly altered to that of finding a technique which would predict failures and then use the lack of prediction of failure as the indicator of successful operation. All of this prediction technique generation needed to be accomplished within the phase of operation of gyro production and system design.

The accumulation of data to be used as a basis for developing failure prediction techniques was done in two ways. A reliability program was established to obtain data on instruments in a controlled manner and to operate those instruments until they were unable to develop angular momentum. This obtained a sufficient amount of data to generate a failure prediction technique. At the same time procedures were established which would collect all field data such that a failure prediction technique could be verified and implemented.

Apollo Gyro

The Apollo gyro, shown in Fig. 1, is a single-degree-of-freedom floated integrating gyro. The gyro wheel is supported by a pair of preloaded angular contact ball bearings. The wheel rotates at 24,000 rpm and developes an angular momentum of 434,000 gm cm^2/sec. The preloading helps to insure a stable location for the wheel center of gravity. A shift in the center of gravity of greater than 1 microinch would result in a change in ADIA drift rate of greater than 7 meru/g. The bearings are adjusted to give an isoelastic structure having equal compliance along the spin and input axes.

The wheel and motor structure are mounted in a hermetically sealed spherical float. This float

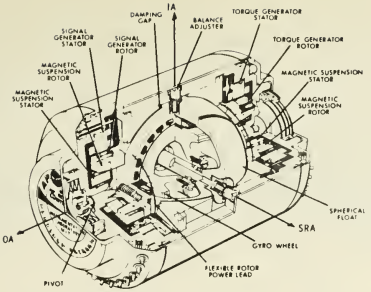


Fig. 1 Apollo IRIG.

is surrounded by a high-density fluid (brominated fluorocarbon), the temperature of which is controlled to maintain the float near neutral buoyancy.

In addition to the buoyancy of the float to minimize the relative motion between the gyro case and float, a magnetic suspension on each end of the unit stabilizes the geometrical relationship between the gyro float and case without causing any resultant torque about the output axis.

The signal generator is a 12-pole multiple E-connected microsyn. Its output is a 3200-cps voltage whose magnitude is proportional to the angular position of the float about the output axis. Its sensitivity is 10 millivolts per milliradian using a 4-volt excitation source.

The torque generator in the Apollo Gyro was designed for digital command angle torquing. A current input to the torque winding causes a torque to the float about the output axis that is proportional to the square of the current input.

The uncommanded torques that act on the gyro float about its output axis can be divided into three types.

1. Those independent of acceleration input. These torques are called bias torques. They are caused by flex-lead restraints and by magnetic restraints due to the magnetic suspensions, signal generator, and torque generator. The drift is termed Bias Drift, BD.

2. Those proportional to acceleration input caused by a noncoincidence of the gyro float center of gravity and center of buoyancy. This term is divided into components along the spin axis and along the input axis. By definition, the center of gravity to center of bouyancy displacement along the spin axis results in the ADIA term (drift due to acceleration along the input axis); that along the input axis in the ADSRA term (drift due to acceleration along the spin reference axis).

3. Those proportional to acceleration input squared. They are caused by nonisoelastic compliance of wheel structure. The compliance terms are A^2DIA, A^2DSRA, $A^2DIASRA$, A^2DIAOA, and $A^2DOASRA$. They cause torques about the gyro output axis for an applied acceleration input. The A^2DIA term causes a torque about the output axis proportional to the square of an acceleration input along the input axis.

The only significant performance terms for the Apollo gyro are BD, ADIA, and ADSRA.

Three of these instruments must operate reliably from a performance standpoint in a system required to navigate three men safely to the moon and back.

The reliability requirements are framed around the lunar landing mission, called the Design Reference Mission (DRM) as shown in Fig. 2. The total time of flight is approximately 200 hours. The gyros operate approximately 10 % of the the total flight time. This time includes the operational time required to initialize the GN&C system and to perform an in-flight or ground alignment of the Inertial Measurement Unit and operation during

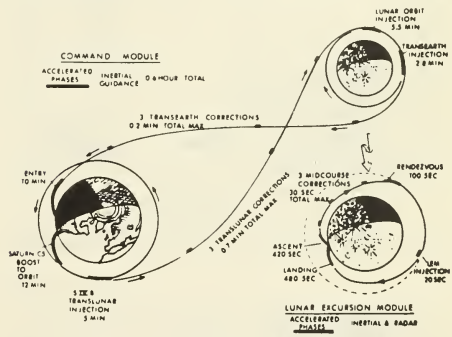


Fig. 2a Guidance mission phases.

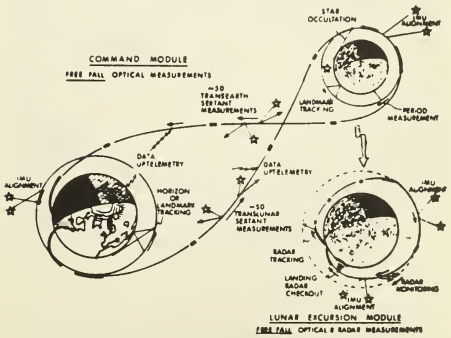


Fig. 2b Navigation mission phases.

thrusting periods. The maximum total operating time including the final ground checkout operations would be somewhat less than 300 hours. The wheel hours per gyro would vary between from 300 hours to about 100 hours depending upon the mode of operation for the gyro wheels. This places the boundary limits on the time of operation for maintenance of angular momentum. The other constraint is that of the degree of performance required to insure successful mission completion. The regions of performance can be seen from Fig. 3 and 4. Figure 3 shows the probability that performance changes of less than plotted amounts would have 'good' mission performance. For example, changes in input axis sensitive drift (ADIA) among any of the three gyros of less than 100 meru/g would not create errors equal to those listed on the figure, and thus present a lower bound on performance requirements. Figure 4 presents the upper bound and shows where performance changes greater than graphed amounts would create serious marginal mission performance. The performance requirements are to have gyro performance deviations in flight lie to the left of the curves in Fig. 3. The margins of safety are to the left of the boundaries in Fig. 4.

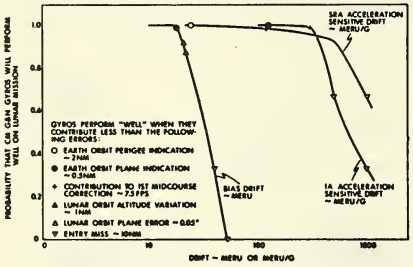


Fig. 3 Mission gyro performance requirements.

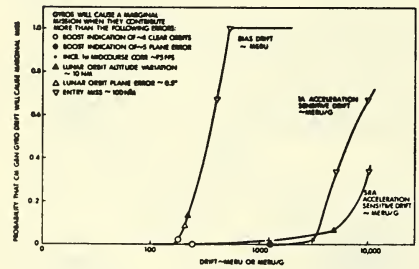


Fig. 4 Marginal mission.

In-flight measurement of gyro performance has been considered. It appears impractical to measure the acceleration sensitive gyro drift terms. It could be done by a position and velocity error comparison between ground and on-board state vectors and then by inferring which specific gyro terms had caused the deviation. It is possible to measure the nonacceleration sensitive gyro drift terms (bias drift) in free fall. After successive IMU alignments had been accomplished, the comparison of the angular errors measured divided by the time interval between measurements will yield the gyro bias drift. It is intended to verify this procedure on one of the early manned flights. The accuracy is limited directly by the quantization of the optical and inertial angle measurements. Thus, the two successive in-flight alignments necessary to have a one sigma uncertainty of 3 meru (0.045 degree/hour) should be made at 2-hour intervals.

System operation is designed utilizing intermittent gyro wheel operation. The question was fundamentally based upon a power conservation consideration. During the DRM of 200 hours the gyro wheels are operated only about 15 % of the time. To operate the gyro wheels for 100 % of the time would increase the fuel cell and battery energy consumption by approximately fifteen kilowatt hours. There would be some other mechanization differences required mainly in the power up sequence for the GN&C system to limit the wheel bearing torques in the gyros.

The mean time before failure (MTBF) requirement for the gyros may now be defined. The gyro operation is approximately 200 hours. This allows for operating time during the last ground checkout, the in-flight operational time, and an adequate margin of safety. Over this time period the gyros must maintain angular momentum and performance changes must be less than defined by Fig. 3 and 4. These performance changes are the limiting boundary. A design and operational goal of 100,000 hours as the MTBF is utilized. This is to say that, for a group of 500 gyros selected good by the performance prediction criterion, it would be expected (with a probability of 67 %) to have not more than one failure in the next 200 hours of wheel operation.

A variety of parameters and test methods are utilized to measure the performance and operation of the gyro unit. The performance tests measure the torques on the float and the float pendulosity about the output axis. Other tests measure parameters of the unit which are performance indicators.

Servo Drift Performance Test

Drift performance data for a gyro instrument are generally taken on a servo loop (Fig. 5). The gyro instrument is mounted on a test table with its input axis parallel to the table axis. The gyro signal generator output is amplified, fed to a table servo loop whose output drives the table torque motor. This rotates the gyro about its input axis. In this loop an ideal gyro would maintain its input axis fixed in inertial space. As viewed from the earth the test table will be rotating at a rate equal to the component of earth rates sensed along the gyro input axis. The table rate for an actual gyro in a servo mode will deviate from earth rate due to torques that act about the gyro output axis. These torques are bias torques independent of acceleration input, torques proportional to acceleration input resulting from non coincidence of the center of gravity and center of

buoyancy of the float, and torques proportional to the square of acceleration input caused by compliance or structural yielding under the acceleration input. The equivalent rate that is required about the gyro input axis to balance out these torques is called gyro drift. They are generally given in terms of a quantity of Earth angular rate. A meru of drift rate is a thousandth of an Earth rate unit (0.015 degree hour).

The GN&C system utilizes a program in the computer to compensate for the expected drift values. They are determined from ground test and are entered into the computer erasable memory during the preflight checkout. Table I is the range of compensation and also the quantization capability.

TABLE I

Gyro Drift Compensation
Block II Gyros

	Magnitude	Quantization
Bias Drift, meru	128	0.1
ADIA, meru/g	630 LM 862 CM	1.0
ADSRA, meru/g	630 LM 862 CM	1.0

CM = Command Module

LM = Lunar Module

Performance errors are, therefore, created when actual performance deviates from expected values.

Since the magnitude of these rates are compensatable, it is the stability of the drift rates that are a measure of the performance of the gyro instrument.

Table rates are measured by determining the time for the table to rotate through prescribed angles (generally one degree is used). Therefore, this type of test measured the average drift rate over a period of time.

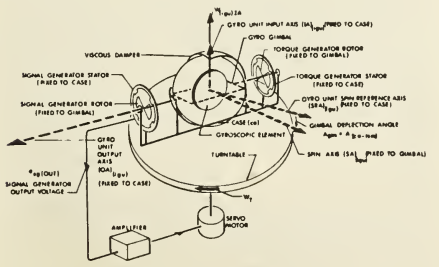
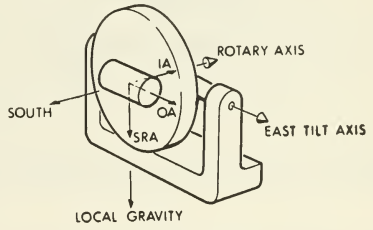


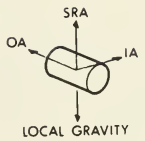
Fig. 5 Line schematic diagram of the Apollo IRIG operating in a single-axis orientational control system.

The magnitude of the components of the drift terms are determined by positioning the table axis so that gravity acts along various gyro axes (Fig. 6). In this way the ADSRA, ADIA, and BD drift terms can be determined by a solution of the three simultaneous equations.

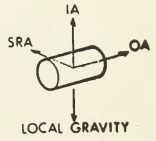
The stability of the individual drift terms are determined by repeated measurements after periods of storage time or by a continuous measurement of table drift rate for a fixed orientation with respect to earth rate.



$\omega_T = \omega_{IE_H}$ - BD-ADSRA



$\omega_T = \omega_{IE_H}$ - BD+ADSRA



$\omega_T = \omega_{IE_V}$ - BD-ADIA

Fig. 6 IRIG calibration.

Torque to Balance

In the torque-to-balance mode the gyro has a closed loop around itself. The signal generator is held at null by torques applied to the torque generator. The torque required in the torque generator is then the sum of the gyroscopic torques, the unbalance torques, and the residual torques. The loop bandwidth is such that torque disturbances less than 3 cycles/minute can be monitored. The changes in current to the torque generator are a measure of the torque disturbance and thus the drift stability. The gyro input axis is generally positioned perpendicular to the earth polar axis such that the total torque measured (torque generator current) is a measure of the total gyro drift. Since the loop does not have to balance components of earth rate, long-term monitoring is possible with an information resolution to less then 0.1 meru.

Rundown Time

When the gyro motor voltage is turned off, the gyro wheel will come to a stop. The wheel speed is braked by the motor magnetic drag, the windage drag, and the frictional torques of the ball bearings.

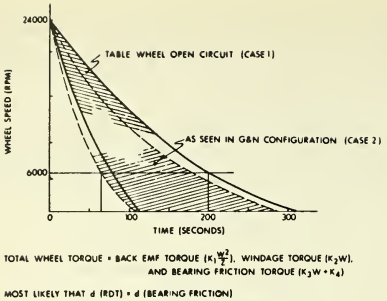


Fig. 7 Wheel speed vs time from turn-off.

Since the motor magnetic and the windage drags are essentially constant for a wheel assembly, the change in rundown time is an indication of the condition of the ball bearings.

Figure 7 illustrates the change in wheel speed plotted against time from wheel voltage turn off, for two cases. Case 1 is for testing on the test table where the wheel is open-circuited soon after the voltage is turned off. Case 2 is for testing in the guidance system where the wheel is loaded by the two other gyro wheel impedances. A load across the wheel shortens the rundown time. There is a large range of expected rundown times for normally operating units. However, the rundown time for a specific unit should not vary by more than 5 seconds from rundown to rundown. Changes in rundown time in a gyro unit are caused by changes in bearing friction. A decrease in rundown time is caused by an increase in bearing friction, while an increase in rundown time is generally caused by a decrease in bearing preload. Rundown time is measured as the time for the wheel speed to decrease from 24,000 rpm to 6,000 rpm and then to standstill. As illustrated in Fig. 7, the total rundown time (24,000 rpm to 0) should give greater resolution as a measurement of the bearing torque, than a rundown time measured to 6,000 rpm.

Retainer Beat Frequency

The rotating ball bearing retainers cause a torque about the gyro output axis at a frequency corresponding to their speed. Since this torque is sinusoidal, its average value is zero and, therefore, causes no drift error. It is detectable at the output of the gyro signal as a sideband of the signal generator frequency (3200 cps) and the retainer frequency (243 cps).

The magnitude of this signal varies periodically due to one retainer lapping or passing the other retainer. A change in this lap frequency is caused by a change in speed of one of the retainers. This frequency stability has been found to be related to the ADIA stability. Figure 8 is a plot of the deviation in ADIA drift data plotted against changes in retainer beat frequency for 25 gyro units. The data for each unit were obtained from twelve hours of servo runs taken with the gyro axis vertically up. Most units fall close to a linear relationship between lap frequency and ADIA stability. Some units do not yield as good a relationship as others. This technique of measuring the ADIA stability by monitoring the stability of a specific frequency on signal generator output is one of the techniques that were developed during the Apollo gyro production program. It is a signal that is available at all stages of spacecraft and system assembly and test and shows promise as a diagnostic performance tool. Additional study of performance evaluation using this technique is still being performed since its simple and easy implementation promises a great advantage over many of the known and presently utilized evaluation methods.

Milliwatt Meter Evaluation

The stability of the power required to operate the gyro motor is an indication of the condition of the gyro wheel bearings. A good wattmeter trace is one having stabilities of 5 milliwatts or about 0.1% of total wheel power. Changes in wheel power are caused by changes in the frictional torques of the wheel bearings. Figure 9 is a typical wattmeter trace taken at various times during the life of a failed gyro instrument. During early hours of its life the power excursions are low, indicating a stable set of wheel bearings. As time progressed the wattmeter trace became more erratic, indicating a decrease in the stability of bearing torques. Finally, the wheel failed to reach synchronous speed and bearing failure had occurred

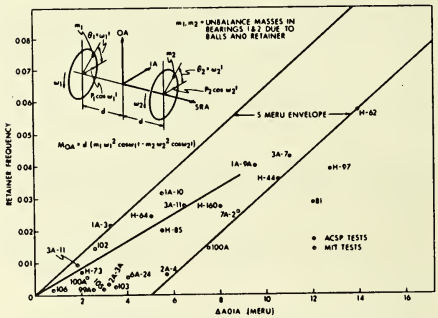


Fig. 8 Change in retainer lap frequency vs ΔADIA for 25 units.

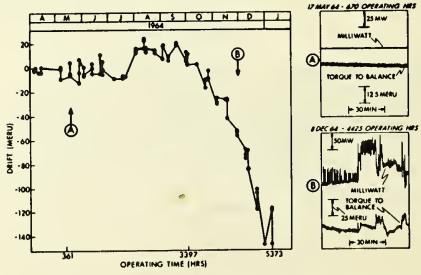


Fig. 9 ADIA drift and wattmeter sample traces vs operating time for Apollo gyro 3A16.

The wattmeter is used as a tool both to screen bearings before they are installed into gyros and to determine if there is bearing deterioration on completed instruments. It is possible to monitor wheel power of gyros installed in their subsystem for Apollo.

Wheel Dynamometer

Before a gyro is built the quality of the wheel bearings can be determined on a dynamometer. The wheel assembly is installed in the dynamometer as shown in Fig. 10. A continuous recording of ball bearing friction is obtained by the feedback loop around the dynamometer gyro bearing shaft.

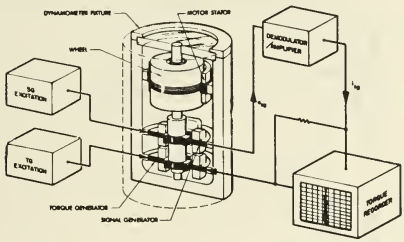


Fig. 10 Ball bearing friction measuring dynamometer.

System Tests

System tests utilized to measure gyro performance fall into two general categories. The first category utilizes angles or changes in angle to evaluate gyro drift. The methods used are to orient the gyro input axis in a direction to have zero drift; then, knowing the gyro orientation with respect to the earth, to calculate the drift performance. There is a second method of utilizing angle information and that is to measure an angular velocity by the use of the gimbal angle transducers in the IMU assuming the base is non-rotating with respect to the earth.

The second and powerful method of measuring gyro performance is to use the rate of change of acceleration as it relates to gyro drift. This is a rapid gyro drift evaluation technique and one which has been automated to a high degree in the Apollo GN&C system for all terms except ADIA. For gyro drift performance measurements including the ADIA term in the spacecraft, a model system is mathematically described. The gravity vector and earth angular velocity vector are calculated and compared to the expected values yielding the vertical drift of the system. An optimum statistical filter separates vehicle motion from the gravity and angular velocity determination.

Reliability Test Program

At the start of the Apollo system design, a gyro reliability test program was initiated. This test program had four aims which were:

1. To obtain a failure model to be used for predicting impending system failures;
2. To determine if running the gyro intermittently, as the Apollo guidance system planned to do to save power, was any less desirable than operating the unit continuously;
3. To determine the effect of vibration on the gyro performance and life; and
4. To determine the relationship between wattmeter data, rundown time, torque to balance, and servo performance as failure indicating parameters.

A total of ten gyros were divided into two groups. Six gyros operated continuously for 200 hours followed by 24 hours of servo performance. Four gyros operated with their wheels on for three hours, off for twelve hours, for a total of 210-hours, followed by 24 hours of servo performance. Except for servo runs, all running hours were accummulated on a torque-to-balance loop.

Wattmeter recordings were taken during the torque-to-balance measurements and rundown time was determined each time the wheels were turned off.

This program lasted for 23 months in which time more than 52,000 wheel hours were accummulated on the ten gyro units. The four wheel bearing failures obtained were used to establish a failure prediction method.

Power required had dictated system operation with the gyro wheels off except when needed for guidance operation. This system operational mode corresponded to the intermittent mode of the reliability test program. Figure 11 is a comparison of the ADIA drift data between those units run continuously and those units run intermittently*. For the same operating time the intermittently operated units resulted in the more stable performance. Therefore, the intermittent mode of operation in the system should not result in a performance penalty.

At the start of the reliability test program, a 3.5 g-rms 20-2000 cps sweep cycle was included at every other performance check. This vibration cycle was discontinued after an accumulation of 30,000 wheel hours had shown no effect on gyro

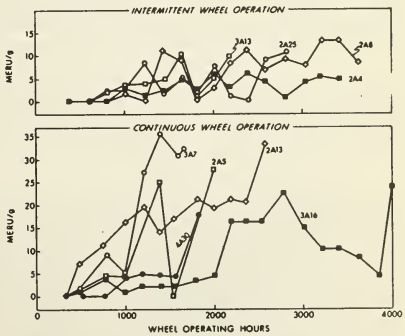


Fig. 11 ADIA drift vs operating time for intermittent and continuous operation.

*There were no differences in the BD and ADSRA terms for the two modes of operating.

performance. Comparing performance data on the gyros when the vibration cycle was in effect to that when it was discontinued again confirmed that this level of vibration did not affect performance.

The relationship between ADIA drift, wattmeter bandwidth, and total wheel rundown time for a gyro that experienced a wheel bearing failure is shown in Fig. 12. A change in rundown time and an increase in wattmeter bandwidth accompanied the ADIA instability and gave warning well in advance of the wheel failure.

Six techniques were considered to lead to a prediction model of gyro performance. These techniques are as follows:

1. Long Term Drift. Gyro drift changes are classified from an initial reference. Variance for each of the three drift terms BD, ADSRA, ADIA are calculated to give the uncertainty, on a population basis, of the drift as a function of time.
2. Autocorrelation. Use of the autocorrelation technique to obtain correlation time of the drift uncertainty of individual gyros as a function of elapsed time or of wheel time.
3. Linear Discriminant. Constants are obtained for which a linear relationship between drift and other gyro parameters can be written.
4. Mean Square Drift Rate. Applied to individual gyro, to calculate variance of the random drift rate process as a function of elapsed time or of wheel time.
5. Probability Transition Matrix. Construct a probability transition matrix from individual gyro drift rate data. The elements of the matrix give the conditional probabilities of the drift rate either staying at the same level or going to another value in a specified time interval. For each time interval a different probability matrix is generated.
6. Physical Model. Devise a model based upon physical aspects of the gyro.

The prediction technique utilized must be one that has a high probability of rejecting bad gyros and yet is reasonably easy to instrument. If at all possible it should utilize parameters which also have a direct bearing on mission performance.

Most of these techniques have been actively analyzed for Apollo Gyros. All but the first method had been eliminated as a performance predicting technique that could be immediately of use for us because of their severe complexity.

From a study of the long-term drift characteristic of the gyro population, it was found that the ADIA drift term for 90 % of the population was stable to better than 25 meru. Since the stability of the ADIA drift term is a measure of the mass stability of the gyro wheel along its spin axis, it can be related to the wheel ball bearing condition. With ball bearing deterioration (the principal failure mode for this instrument) a technique related to the ADIA drift term obtained in the reliability test program appeared to be the most promising predicting device. This model, called the Delta 25 criterion, will now be discussed.

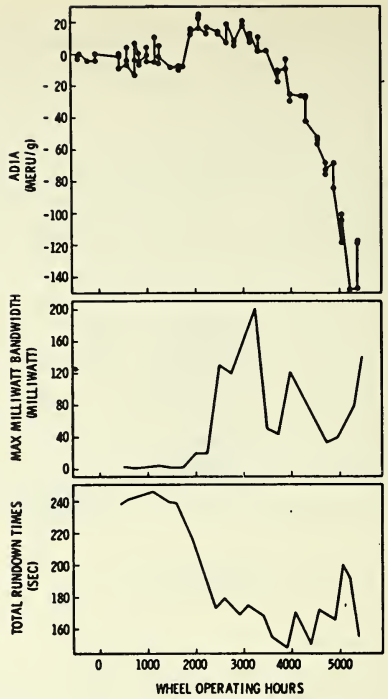


Fig. 12 ADIA drift, wattmeter bandwidth, and rundown time vs operating time for Apollo gyro 3A16.

The Delta 25 Criterion

Any gyro unit that exhibits a shift in drift terms of more than 25 meru or 25 meru/g is suspect and may not be used for flight. To meet the criterion this shift of 25 meru/g must occur at a single test location, must be verified by a second data point, and must have occurred across a storage period of less than four months. The single test location eliminates any data shifts caused by test equipment variability. The verification by a second data point eliminates bad data. The restriction on storage period between tests eliminates storage sensitive shifts that have been noted in this instrument.

Figure 13 is a plot of unit 3A16 whose bearings failed in the reliability program. At approximately 2100 wheel hours of operation the ADIA drift data changed by more than 25 meru/g. The gyro was in the same location, and the performance change was verified by a second test.

At this point, it was predicted that this gyro would fail. The gyro operation was continued in the test program in order to verify the performance throughout its life. For the next 1500 hours the performance of the gyro was very respectable in that the total acceleration sensitive drift along the input axis was less than 30 meru/g. At about 3600

hours of wheel time, the performance changed again by more than 25 meru/g and following this performance change there was some 1800 hours of wheel time before the gyro failed completely. After each Delta 25 indication the total change in ADIA for the next 300 hours was less than 28 meru/g. There was a large lead time following a change in performance before the unit failed; during this period the performance was still suitable for an Apollo Mission. The performance of the other two drift terms, the bias drift and the acceleration sensitive drift along the spin reference axis, did not change significantly during the entire operation of the instrument from beginning to end. This is characteristic of all wheel failures.

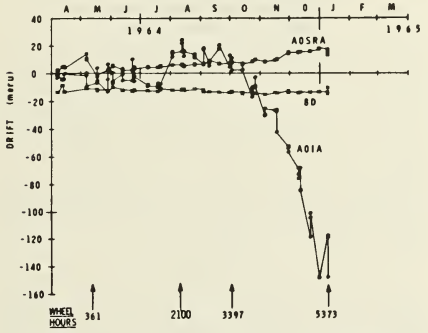


Fig. 13 25 IRIG 3A16 drift coefficients.

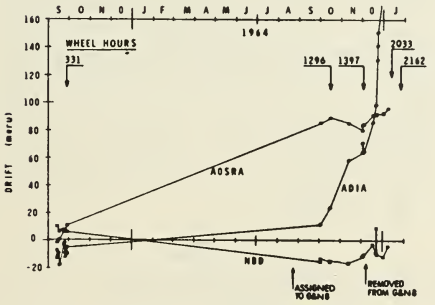


Fig. 14 25 IRIG 2A11 drift coefficients.

Gyro 2A11 as shown on Fig. 14 is an example of prediction of system performance where the gyro had been installed in the guidance system. The ADSRA drift term of this gyro changed by more than 25 meru/g at the time of installation into the system. Subsequent tests verified that this change had occurred. However, the Delta 25 criterion would not have picked up this parameter on that instrument which had changed location. After installation in the system a subsequent change of more than 25 meru/g occurred in ADIA and was verified.

At that time this instrument was recognized as a potential failure and was removed from the guidance system. The gyro was subsequently tested and operated many times in the OFF/ON mode until the wheel finally failed. Following detection of the change in performance by more than 25 meru/g, there was an accumulated time of operation equivalent to 80 lunar missions. Again, the performance of the instrument remained stable for the two terms, ADSRA and bias drift, until the gyro failed completely. The other instruments in which wheel failures occurred also indicated a long lead time from the time of change in performance until the wheel failed completely. The Delta 25 test is now utilized as a performance predictor for all gyros.

To fully implement the performance prediction criterion the performance of the gyros are periodically measured from the time the units are accepted until they are flown in guidance systems.

After the gyros pass an acceptance test, the gyros are next tested in a retest program. This program measures the performance of all units not installed in systems. From the retest data, units are selected for system use. Units whose retest data are not adequate for system assignment are re-routed for a second retest, repaired, or assigned as laboratory test units.

After the units are assigned and assembled into systems, gyro performance in an inertial subsystem are measured. These performance tests are taken before and after a system vibration. Then gyro performance is measured in a completed guidance system and later in the spacecraft.

At all times performance data are taken, and are compared to the failure criterion. Units that exceed the criterion are removed from the system to be verified as a failure or are requalified. The requalification test is a re-run of the unit acceptance test.

Usefulness and Efficiency

The objective of failure prediction is to be able to predict at least 200 wheel hours with good performance, thus assuring gyro operation and system performance for the lunar landing mission. The usefulness of the failure prediction method can be characterized by the proportion of bad gyros remaining after failure prediction.

The usefulness index is defined in Fig. 15. The calculated usefulness index for the Apollo I gyro is 99% and is 98% for the Apollo II Gyro. The efficiency of the method can be measured by the

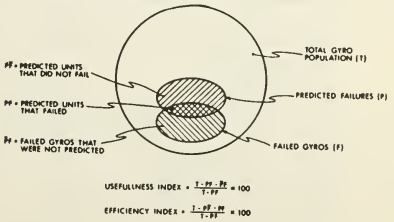


Fig. 15 Usefulness and efficiency.

67-588

proportion of good gyros rejected by the prediction technique. The efficiency index is defined in Fig. 15. The calculated efficiency for the Apollo I gyros is 89% and 98% for the Apollo II gyro.

To apply any performance prediction technique it is necessary to have sufficient information upon which to base any decision. Therefore, at the onset of the program, a procedure was established for the central collection and assessment of all gyro and accelerometer data. This data collection flowgram is shown in Fig. 16. The checkout procedures were written so that a copy of all inertial performance data was sent to MIT. The flowgram or data from a single system is shown in Fig. 17. This is the characteristic data flow for a single gyro starting at acceptance test and following it through flight. For a single cycle, one parameter of one gyro would be measured twenty-two times. The Z gyro used in the SA 202 mission was measured forty-four times. This is the established flow of performance data.

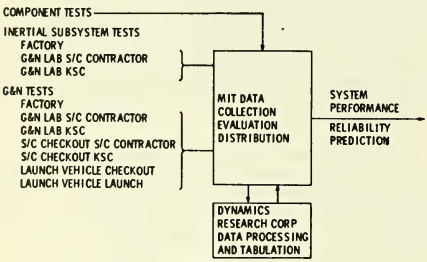
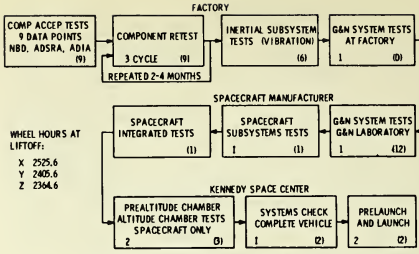


Fig. 16 Data collection flow.



S/C 011 Z GYRO DATA TAKEN INDICATED BY NUMBERS IN PARENTHESIS. OTHER NUMBERS REPRESENT NORMAL PERFORMANCE EVALUATION DATA POINTS.

Fig. 17 Inertial component performance - data sequence.

The second portion of this was to mechanize in both the system and ground support equipment the ability to measure the performance parameters. At the highest level of assembly, the GN&C system in the spacecraft stacked on the pad and under sway conditions, the two acceleration sensitive drift terms and the bias drift are measurable. It is possible by removing connectors to measure other performance indicators such as rundown time. The self-contained test methods used in the spacecraft will measure drift performance with an error in measurement of less than 5 meru/g.

The data upon collection are converted to punch cards and a magnetic tape in a format suitable for computer processing and printing. A monthly listing is printed both as an inventory and as a data tabulation. An example of the format used is shown in Fig. 18. It tabulates for each instrument thirteen quantities unique to that instrument in addition to the performance and location history. The rapid access to all gyro data has made possible a number of analyses. For example, long-term drift can be examined easily and quickly. All data plots used here are automatically made from the gyro data file.

NASA 3A-24 ACE AE-119-3

DA MO YR	LOC	TYP	ASSN	WHEEL HOURS	NBD	ADSRA	AOIA	ID RDT	DELSF+	DELSF-	I+	I-	NOTES
2/FE/65	A03	SPR	17Z	660					771	266	114.942		
2/FE/65	A03	SPR	17Z	690 -	8.1	1.7							
2/FE/65	A03	SPR	17Z	693 -	7.0	- 1.9			572 -	472	114.942		
14/FE/65	A03	SPO	17Z	840 -	8.7	0.0							
14/FE/65	A03	SPO	17Z	849 -	7.0	- 1.1			559 -	1017	114.993		
7/MR/65	A03	SVM	17Z	995 -	6.6	0.6 -							DQI
7/MR/65	A03	SVM	17Z	1000 -	9.4	- 0.3			160 -	376	114.952		
19/MR/65	N02	G&N	17Z	1114					- 1020		114.954	114.936	
20/MR/65	N02	G&N	17Z	1134 -	6.9	1.1							
23/MR/65	N02	G&N	17Z	1167									RUN2
23/MR/65	N02	G&N	17Z	1174									POS1 ONLY
											9/11/65-ABORT	244	
12/MY/65	N02	G&N	17Z	1237					- 890		114.937	114.937	NAV DSKY PROBLEM
					13MY65 AGE 28 PWR SUPPLY RMV FROM TS1 & INSTALLED IN TS2								
19/MY/65	N02	G&N	17Z	1290 -	4.7	0.3 -	0.2						1A FROM SUMMARY
26/MY/65	N02	SCK	17Z	1359 -	10.5 -	1.9							
26/MY/65	N02	SCK	17Z	1362 -	4.9	- 9.3			- 535 -	306	114.993	114.958	
24/JE/65	N06	G&N	17Z	1407 -	7.6	1.2							
											29JE65 INVESTIGATION PERFORMED		
29/JE/65	N06	G&N	17Z	1463 -	7.2	1.6							27.9VDC
								TEST AFTER ABORT 10080 ON MOTHERS DIRECTION					
30/JE/65	N06	G&N	17Z	1466 -	7.7 -	2.3							29.0VDC
30/JE/65	N06	G&N	17Z	1478 -	6.8	0.6							28.8VDC
1/JL/65	N06	G&N	17Z		6.2	1.3							27.9VDC
											VDC>27.9, INST. AT 23.9		
7/JL/65	N06	G&N	17Z						- 1861		114.947	114.937	SUNRISE 45 AVG SF
7/JL/65	N06	G&N	17Z						244	1770			SUNRISE 69
8/JL/65	N06	G&N	17Z		4.7	4.0 -	7.2		244	1831			SUNRISE 69
30/AU/65	N02	G&N	17Z	1558 -	6.6	0.6 -	0.3						ETR-039
2/SE/65	N02	G&N	17Z	1580 -	10.1	9.3 -	6.9						ETR-039
				3SE65 TO N06, 22SE65 RMV FROM N06 IN PREPARATION OF S/C INSTALLATION									
27/FE/66	N3	G&N	17Z	1722 -	9.4 -	0.4			1450 -	748			S/C 011
21/MR/66	N	G&N	17Z	1762					2329	249			S/C 011
27/MR/66	N	G&N	17Z	1893 -	5.8	0.4							S/C 011
26/MY/66	K10	G&N	17Z	2014 -	3.1 -	3.2				- 793			
26/MY/66	K1D	G&N	17Z	2025 -	4.0 -	4.6							
27/MY/66	K10	G&N	17Z	2036 -	0.1	9.7							
2/JE/66	K10	G&N	17Z	2055					1646 -	917			
11/JL/66	K1D	G&N	17Z		6.4 -	3.9			1626 -	340			
5/AU/66	K1D	G&N	17Z		6.4	3.2				- 799			
23/AU/66	K	G&N	17Z		9.2	0.0							
29/AU/66				2344									· Liftoff

Fig. 18 Sample monthly inventory and data tabulation.

The output of the entire program is the reliability prediction which is utilized for each gyro.

The F Criterion

Studies of wheel bearing failures have shown a correlation between ADIA instability and an impending wheel failure. For this reason any parameter that measured ADIA stability could be used as a prediction of an impending bearing failure.

Three statistical parameters which could be applied to the ADIA drift data are:

1. A magnitude change in ADIA drift exceeding some level;
2. The standard deviation of the ADIA drift exceeding some level; and
3. The RMSSD (root mean squared value of successive differences) exceeding some level.

Method 1 above is the basis of the Delta 25 criterion that was previously discussed.

All methods are found to obtain large magnitudes for a situation that occurs frequently but is not an incipient failure. For example, changes in drift can occur due to dimensional changes in the float structure. This results in a magnitude change in drift that is found to stabilize at a new level. It is difficult to have a standard magnitude or deviation criterion that will account for a stable level change. However, it can be implemented easily in an RMSSD (root mean squared sum of differences) criterion. This method, called the F criterion, states that a failure is predicted if the RMSSD of the last eleven ADIA data points exceed 10 meru/g. A unit can requalify if its RMSSD later drops below 10 meru/g. The requalification indicates a unit that shifted in ADIA drift but later stabilized.

Figure 19 is a plot of the RMSSD of unit 2A5 plotted against wheel hours. This instrument met the F criterion for failure prediction at approximately 1500 wheel hours. Five hundred hours after failure prediction the wheel failed to start at the application of wheel voltage. These hours would have been sufficient time to complete two pre-flight tests and missions.

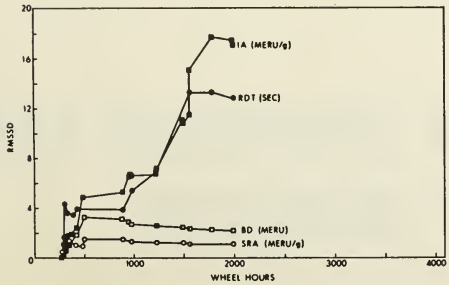


Fig. 19 RMSSD vs wheel hours for Apollo gyro 2A5.

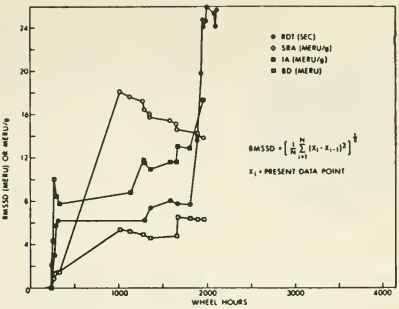


Fig. 20 RMSSD vs wheel hours for Apollo gyro 2A11.

The RMSSD drift terms are plotted vs. wheel hours for gyro 2A11 in Fig. 20. This unit met the F criterion during acceptance testing (approximately 300 wheel hours) when the RMSSD of the ADIA exceeded 10 meru/g but later requalified itself when the RMSSD of the ADIA drift dropped below 10. This unit then exceeded the F criterion at approximately 1200 hours. The unit had stable ADSRA and NBD drift performance even after it had exceeded the F criterion for ADIA data. The gyro continued to run in an OFF/ON mode for the next 800 hours. In this example, as in the previous example, there was adequate warning of an impending failure to allow a successful completion of the mission.

When the F criterion was applied to all gyros in the Apollo program, all but one of the Apollo I bearing failures were predicted and 23 units whose failure had been predicted have not failed yet. This yields a usefulness index of 99 % and an efficiency index of 80 %.

For the Apollo II gyro the F criterion predicted all three of the bearing failures, for a usefulness index of 99 %. It also predicted eleven units that have not yet failed, for an efficiency index of 96 %.

Mean Time Before Failure of Gyros in Production

Table II is a summary of the Apollo Gyro program. As of January 1, 1967 the 361 Apollo production gyros have run a total of 380,771 wheel hours. Excluding the wheel hours on any units after they were failure predicted by the Delta 25 criterion the total was 346,175 hours. Of the remainder, two units were not removed from systems by the criterion and would be considered as flight-critical failures. These two failures in 346,175 operating hours yield a mean time before failure of 100,000 hours with 70 % confidence.

Conclusion

This paper discussed the Apollo gyro history. It was necessary to make an order-of-magnitude increase in the reliability of this instrument when used for the Apollo mission. To accomplish this objective it was necessary to test the gyro many times before it would be flown in the system. Testing will yield some failures but it was necessary that this testing would also give evidence of impending

failures. No failure predicting technique can be expected to give forewarning that a solder connection will open or that the gyro float will come into contact with a contaminant in the damping fluid. Only careful workmanship can make this type of failure a minor contributor to gyro reliability.

TABLE II

	APOLLO I	APOLLO II
Production Units		
Program Start	March 1962	August 1964
First Unit	February 1963	May 1965
Program End	June 1965	February 1967
Total Quantity	126	235

	Number Built	Total Wheel Hours
Production Units	361	380,771
MIT-Built Units	22	61,666
10 APOLLO I		
12 APOLLO II		
Total	383	442,437

A failure model was needed at the start of the program. This was implemented by a reliability test program run to accummulate wheel hours and failure under controlled test conditions. The failure mechanisms for the Apollo gyro are quite complex. All units do not fail in the same way. A study of failed gyros obtained a criterion that was common to all the failures and was applied as the Delta 25 failure criterion. This technique is not expected to be 100% effective as it will predict some units that are not about to fail and, conversely, it will miss some units that do fail. It requires elimination of data that occurred across location changes even though data have been shown to correlate very well at all locations. The application of this criterion since it is empirical had in its favor that it worked. As more data are accummulated more refinement can be made in a prediction criterion. This is being done and the F criterion is one of the possible refinements. The overall program is such that extensive data cannot always be obtained for all predicted failures. Since the techniques as now used do work they were presented in this paper as a good starting point for failure prediction.

ACKNOWLEDGEMENT

This report was prepared under DSR Project 55-23850, sponsored by the Manned Spacecraft Center of the National Aeronautics and Space Administration through Contract NAS 9-4065.

The publication of this report does not constitute approval by the National Aeronautics and Space Administration of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

BIBLIOGRAPHY

1. Palmer, Peter J., Gyro Torque Coefficients, MIT/IL Report No. E-1601, Revision 3, Cambridge, Mass., January 1966.

2. Bonafede, William H., Gyro evaluation by the Table Servo Method, MIT/IL Report No. E-1509, Cambridge, Mass., February 1964.

3. Hall, E. J., Miller, J. E., Aronson, J., Specifications for Procurement of Apollo Inertial Reference Integrating Gyro, MIT/IL Report R-348, Cambridge, Mass., December 1961.

4. Draper, C. S., Wrigley, W., Hoag, D. G., Battin, R. H., Miller, J. E., Koso, A., Hopkins, A. L., and Vander Velde, W. E., "Space Navigation Guidance and Control," Technivision Limited, Braywick House, Maidenhead, Berkshire, England, 1966.

5. Schmidt, G. T., and Brock, L. D. Statistical Estimation in Inertial Navigation Systems, AIAA Guidance and Control Conference, 1966, Seattle, Wash.

No. 67-589

A DESCRIPTION OF THE CMG AND ITS APPLICATION TO SPACE VEHICLE CONTROL

by

B. J. O'CONNOR and L. A. MORINE
The Bendix Corporation
Teterboro, New Jersey

AIAA Paper
No. 67-589

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

-- NOTES --

A DESCRIPTION OF THE CMG AND ITS APPLICATION TO SPACE VEHICLE CONTROL

B. J. O'Connor, Associate Fellow, Assistant General Manager
L. A. Morine, Senior Engineer
Navigation and Control Division
The Bendix Corporation, Teterboro, N. J.

ABSTRACT

High pointing accuracies and long time duration missions of large spacecraft, such as the Apollo Telescope Mount (ATM) make control systems utilizing momentum exchange devices attractive. Momentum exchange devices are desirable for the following reasons: 1) continuous control, 2) recoverable energy source, 3) efficient control of cyclic disturbances, 4) ease of propellant management. The control moment gyros are particularly advantageous over reaction wheels for the control of large spacecraft because of: 1) better efficiency, 2) larger output moments, 3) better bandwidth and dynamic range, and 4) greater range of linearity. The principle of momentum exchange, a description of the control moment gyro (CMG), and application to spacecraft control are presented. Particular emphasis is placed on usable CMG control laws for the ATM control system along with their corresponding methods of implementation. A derivation of the CMG kinematic equations of motion is outlined. These equations are then combined with the CMG servo characteristics to describe the rigid body dynamics of the ATM control moment gyro.

1. INTRODUCTION

This paper will describe a control moment gyro (CMG) and its use in a space vehicle attitude control system. The discussion is divided into two basic parts.

A) A discussion of an attitude control system with CMG actuators and possible control laws which will convert attitude error signals to desired reaction moments on the vehicle

B) A description of the dynamic characteristics of a CMG.

1.1 MOMENTUM EXCHANGE

The principle of momentum exchange is based on Newton's second law of angular motion which states that the total external moment, $\bar{M}_{EXT}$, acting on a system, is proportional to the time rate of change of angular momentum with respect to inertial space and can be written as

$$\bar{M}_{EXT} = \left\{ \frac{d\bar{H}\ \text{system}}{dt} \right\}_{\text{I space}} \qquad (1.1)$$

If the system consists of two momenta, $\bar{H}_{vehicle}$ and $\bar{H}_{control}$ then integrating equation (1.1) yields

$$\int \bar{M}_{EXT}\, dt = \bar{H}_{EXT} = \bar{H}_{control} + \bar{H}_{vehicle} - \bar{H}_{system}^{(0+)}, \qquad (1.2)$$

which provides the equation defining the principle of momentum exchange. If the controller does not expel mass, then changes in controller momentum can be used to balance external torques as well as change the spacecraft attitude by varying the vehicle angular momentum. Equation (1.2) states that it is possible to transfer or exchange momentum between the controller and the vehicle.

Momentum exchange is desirable for the following reasons:

A) Continuous control

B) A recoverable energy source

C) Efficient control of cyclic disturbances

D) Ease in management of propellant expulsion for continuous disturbances.

Continuous control and a recoverable energy source are the result of the use of electrical energy as the prime source of power. From equation (1.2) it can be seen that the momentum exchange device can handle cyclic torques on a continuous basis over long time periods. Constant external moments applied to the vehicle will cause the controller to eventually reach its maximum capacity, and thus saturate the momentum exchange device. This saturated condition requires the expulsion of propellants to remove some momentum from the momentum device; however, the propellant expulsion can be accomplished at convenient times. Hence propellant expulsion management can be accomplished in a more sophisticated manner. Other external moments may be used to advantage in some cases such as gravity gradient, magnetic moments, etc.

Momentum control is presently implemented by the use of the following basic devices:

A) Reaction wheels

B) Control moment gyros.

A reaction wheel has a high inertia rotor that can be accelerated by an electromagnetic motor. The torque which accelerates the wheel, and causes a change in the magnitude of the device's angular

67-589

momentum, reacts on the vehicle through the motor electromagnetic field.

The CMG is basically a gimballed wheel rotating at a constant speed which provides a constant angular momentum magnitude capable of variable orientation relative to the spacecraft. A moment is imparted to the vehicle by causing a change in direction of the constant momentum magnitude. A single constant momentum magnitude device is not usable since equation (1.2) requires a change in the magnitude as well as direction of the vector $\bar{H}_{control}$. Thus two or more CMGs must be used to effect both a magnitude and orientation change of $\bar{H}_{control}$.

As can be seen from Figure 1.1, if vectors $\bar{H}_{(1)}$ and $\bar{H}_{(2)}$ are each constant in magnitude and both have two degrees of freedom, it is possible to control the magnitude of the momentum vector in three degrees of freedom.

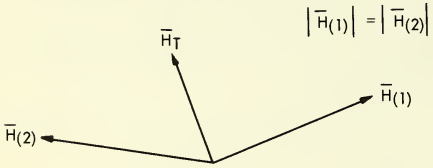


FIGURE 1.1 VARIABLE MOMENTUM EFFECT OF TWO CMGs

The maximum momentum possible in the system is obtained when all CMG momentum vectors are parallel and thus momentum exchange is possible for all CMG cluster momenta within a momentum volume as shown in Figure 1.2. The shape of this volume depends on the particular CMG cluster configuration and the condition of saturation is the surface of this volume, thus a momentum magnitude outside of this volume is unattainable.

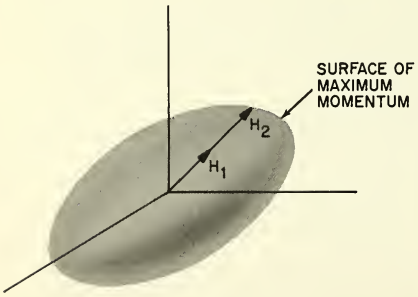


FIGURE 1.2 MOMENTUM VOLUME

1.2 THE ADVANTAGES OF A CMG

A cluster of CMGs offers significant advantages over a reaction wheel when large maximum moments are required for large vehicle control.

These advantages are:

A) Better efficiency

B) Larger maximum moments and an increased dynamic range of moments

C) Better bandwidth characteristics

D) A more linear approach toward saturation.

The CMG efficiency is greatly improved over a reaction wheel for general operating momentum values since the CMG wheel operates at one speed for which the efficiency can be optimized.

Larger maximum moment control is easily obtained with modest turning rates of the large constant momentum. Without great complexity, torque ranges of 2000:1 can be obtained.

Better bandwidth characteristics of the CMG are obtained from an angular velocity type gimbal servo. The reaction wheel rotor time constant is large due to its physical characteristics and thus limits its bandwidth.

A linear approach toward saturation is possible with the CMG since the gimbal control rates never approach a saturated state. The reaction wheel as it approaches its maximum rate tends to become non-linear.

1.3 A DESCRIPTION OF A CMG

The CMG is a two degree of freedom gyroscopic device and is shown schematically in Figure 1.3. It consists of a constant speed wheel held in a

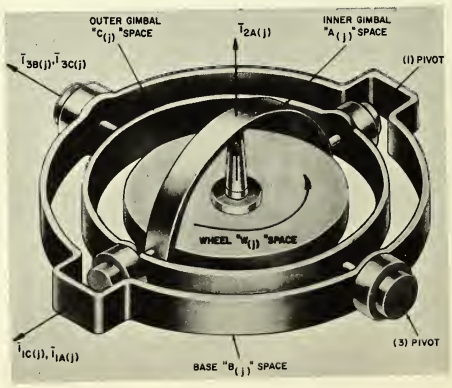


FIGURE 1.3 THE "j" CMG SCHEMATIC

housing which is called the inner gimbal. The inner gimbal is coupled to the outer gimbal through the (1) pivot which is perpendicular to the wheel spin vector. The outer gimbal is held to the base through the (3) pivot which is perpendicular to the (1) pivot. The (1) and (3) pivots are driven by geared D.C. motor torquers.

There are three spaces used to describe the equations of motion of the jth CMG as shown in Figure 1.3.

A) "$A_{(j)}$" space or inner gimbal which is described by the coordinate system $\bar{1}_{1A(j)}$, $\bar{1}_{2A(j)}$, $\bar{1}_{3A(j)}$

B) "$C_{(j)}$" space or outer gimbal which is described by the coordinate system $\bar{1}_{1C(j)}$, $\bar{1}_{2C(j)}$, $\bar{1}_{3C(j)}$

C) "B(j)" space or the base of the CMG which is described by the coordinate system $\bar{1}_{1B(j)}$, $\bar{1}_{2B(j)}$, $\bar{1}_{3B(j)}$.

The "$A_{(j)}$" space coordinate system fixed to the inner gimbal is defined such that the $\bar{1}_{2A(j)}$ vector is along the SRA axis, the $\bar{1}_{1A(j)}$ vector is along the (1) or inner pivot and the $\bar{1}_{3A(j)}$ vector is the third vector in the orthogonal coordinate system.

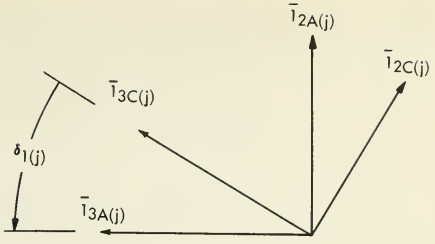
The "$C_{(j)}$" space coordinate system fixed to the outer gimbal is defined such that $\bar{1}_{1A(j)}$ and $\bar{1}_{1C(j)}$ are coincident and the $\bar{1}_{3C(j)}$ vector lies along the (3) or outer pivot. The $\bar{1}_{2C(j)}$ vector is the remaining vector in the orthogonal coordinate system.

The zero position of the CMG is defined when the vectors $\bar{1}_{1A(j)}$, $\bar{1}_{2A(j)}$ and $\bar{1}_{3A(j)}$ are coincident with vectors $\bar{1}_{1C(j)}$, $\bar{1}_{2C(j)}$, $\bar{1}_{3C(j)}$ and $\bar{1}_{1B(j)}$, $\bar{1}_{2B(j)}$ and $\bar{1}_{3B(j)}$, respectively.

The "$B_{(j)}$" space coordinate system fixed to the base is defined such that the vectors $\bar{1}_{3C(j)}$ and $\bar{1}_{3B(j)}$ are coincident. The $\bar{1}_{2B(j)}$ vector normally points toward the top of the CMG.

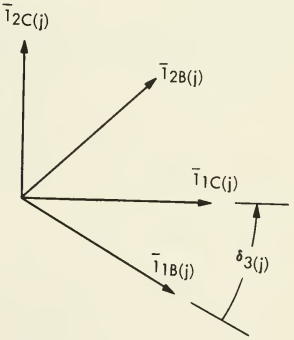
The (1) pivot angle, $\delta_{1(j)}$, is positive when the inner gimbal is rotated in a positive direction about the $\bar{1}_{1A(j)}$ vector with respect to the outer gimbal, as shown in Figure 1.4.

The (3) pivot angle, $\delta_{3(j)}$, is positive when the outer gimbal is rotated in a positive direction about the $\bar{1}_{3C(j)}$ vector with respect to the base as shown in Figure 1.5.



NOTE:
THE $\bar{1}_{1A(j)}$, $\bar{1}_{1C(j)}$ VECTORS POINT OUT OF PAPER.

FIGURE 1.4 DEFINITION OF (1) PIVOT ANGLE $\delta_{1(j)}$



NOTE:
THE $\bar{1}_{3C(j)}$, $\bar{1}_{3B(j)}$ VECTORS POINT OUT OF PAPER.

FIGURE 1.5 DEFINITION OF (3) PIVOT ANGLE $\delta_{3(j)}$

The transformation matrices from "$A_{(j)}$" to "B(j)" space, "A(j)" to "$C_{(j)}$" space and "$C_{(j)}$" to "B(j)" space can be obtained from the definition of coordinate systems and gimbal angles and are given in equations (1.3) through (1.5).

$$\begin{pmatrix} H_{1B(j)} \\ H_{2B(j)} \\ H_{3B(j)} \end{pmatrix} = \begin{pmatrix} \cos\delta_{3(j)} \begin{bmatrix} -\sin\delta_{3(j)} & \cos\delta_{1(j)} \end{bmatrix} & \begin{bmatrix} \sin\delta_{3(j)} & \sin\delta_{1(j)} \end{bmatrix} \\ \sin\delta_{3(j)} \begin{bmatrix} \cos\delta_{3(j)} & \cos\delta_{1(j)} \end{bmatrix} & \begin{bmatrix} -\cos\delta_{3(j)} & \sin\delta_{1(j)} \end{bmatrix} \\ 0 \qquad\qquad \sin\delta_{1(j)} & \cos\delta_{1(j)} \end{pmatrix} \times \begin{pmatrix} H_{1A(j)} \\ H_{2A(j)} \\ H_{3A(j)} \end{pmatrix} \qquad (1.3)$$

$$\begin{pmatrix} H_{1C(j)} \\ H_{2C(j)} \\ H_{3C(j)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\delta_{1(j)} & -\sin\delta_{1(j)} \\ 0 & \sin\delta_{1(j)} & \cos\delta_{1(j)} \end{pmatrix} X \begin{pmatrix} H_{1A(j)} \\ H_{2A(j)} \\ H_{3A(j)} \end{pmatrix} \quad (1.4)$$

$$\begin{pmatrix} H_{1B(j)} \\ H_{2B(j)} \\ H_{3B(j)} \end{pmatrix} = \begin{pmatrix} \cos\delta_{3(j)} & -\sin\delta_{3(j)} & 0 \\ \sin\delta_{3(j)} & \cos\delta_{3(j)} & 0 \\ 0 & 0 & 1 \end{pmatrix} X \begin{pmatrix} H_{1C(j)} \\ H_{2C(j)} \\ H_{3C(j)} \end{pmatrix} \quad (1.5)$$

The use of these transformation matrices is implied although not specifically referred to in sections 3 through 6 of this paper.

The gimbal attitudes will be controlled by angular velocity servos as shown schematically in Figure 1.6. Each pivot has a rate generator which is directly coupled to the D.C. motor and senses the relative gimbal rates $\dot{\delta}_{1(j)}$ and $\dot{\delta}_{3(j)}$.

The servo amplifier drives the D.C. motor with a current source and has a lag network to increase the static gain in the loop. The function G(p) is defined in equation (1.6).

$$G(p) = K_{WT}\frac{(\tau_2 p + 1)}{(\tau_1 p + 1)}, \quad \tau_1 > \tau_2 \quad (1.6)$$

Throughout the discussion contained in this report, the CMG will always be considered as described in this section when reading sections 3 through 6. In section 2, a single CMG is discussed and it can be thought of as a two degree of freedom unit with its outer gimbal locked to the base. When a single CMG is discussed it will be abbreviated as SCMG and in all cases the double CMG will be abbreviated as CMG.

The (j) subscript was used in definition of spaces and gimbal angles since section 3 and 4 consider a cluster of three CMGs. In sections 5 and 6 the (j) subscript was omitted since there was no need to identify the CMG orientation as shown in Figure 2.8.

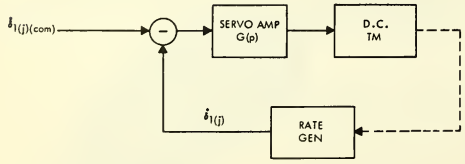


FIGURE 1.6 THE (1) PIVOT ANGULAR VELOCITY SERVO

2. AN ATTITUDE CONTROL SYSTEM WITH CMG ACTUATORS

An attitude control system with CMG actuators will be described in a standard or typical configuration and this configuration basically defines the Apollo Telescope Mount (ATM) Control System.

The standard configuration uses CMGs without undo complications and is basically a position servo with attitude rate stabilization. Some of the major functions of the control system are:

A) Position sensor

B) Vehicle control law

C) CMG control law

D) CMG configuration

E) Vehicle dynamics

F) Momentum management.

2.1 SYSTEM DESCRIPTION

A block diagram of the standard control system is shown in Figure 2.1.

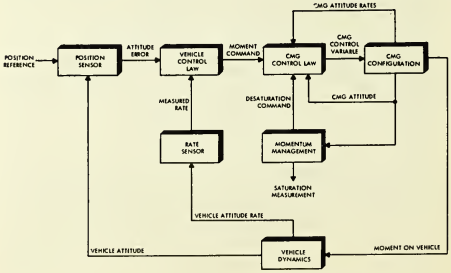


FIGURE 2.1 VEHICLE ATTITUDE CONTROL WITH CMGs

The position sensor which is an optical device will measure the attitude error between the actual vehicle attitude and the desired attitude and this error is combined with the vehicle angular rate in the vehicle control law. The vehicle control law is basically a proportional plus differential controller which determines the desired command moment, $\bar{M}_{com}$, to be applied to the vehicle.

The command moment is processed by the CMG control law and the outputs of this law provide the input commands to the CMG gimbal servos.

The command variable to the CMG servos drives the individual momentum vectors of each CMG and this change in angular momentum causes a reaction moment on the vehicle.

The reaction moments on the vehicle are forcing functions which act on the vehicle dynamics and change the vehicle attitude and attitude rate thus closing the control loop.

2.2 THE ATTITUDE CONTROL SYSTEM PROPERTIES

The properties of each block in any control system depend naturally on the particular control system application; however, there are some general properties of these elements which will define a standard system. With the exception of a CMG control law and CMG configuration, the remainder of the blocks in Figure 2.1 are common to most control systems. These general properties are well known and are mentioned briefly as follows:

A) Position Sensor - measures the attitude error between the actual vehicle attitude and the desired vehicle attitude.

B) Vehicle Control Law - formulates the vehicle moment command (desired vehicle moment) as a function of vehicle state variables and their derivatives and/or integrals.

C) Vehicle Dynamics - the angular equations of motion of the vehicle with controller moments and external moments as forcing functions.

The system functions that involve the CMG are contained in two categories:

A) Control loop functions

B) Momentum management functions.

The CMG control loop functions involve the CMG control law and the CMG configuration in such a manner as to affect desired command moments on the vehicle. The momentum management involves the saturation measurement, criteria for desaturation, and the actual mechanization for desaturation.

CMG Control Loop Functions

The individual properties of the CMG control law and CMG configurations can vary radically depending on the application. Generally, the desired total transfer functions for both CMG blocks in Figure 2.1 satisfy the identity given in equation (2.1).

$$\bar{M}_{command} = \bar{M}_{vehicle} \qquad (2.1)$$

Physically this is accomplished by the cross moments from one CMG being cancelled by another CMG in the cluster configuration with a net resultant moment equal to the command moment. Complete cancellation of the cross moments is not always practical; however, since many other factors or systems constraints enter into the choice of CMG configuration and the control law, some degree of cross coupling must be tolerated. Some of the more important factors are:

A) Size, weight and power

B) Simplicity of control law (major cause of cross coupling)

C) Momentum utilization (a function of expected moment disturbances)

D) Degree of redundancy required by the system

E) The number of axes to be controlled

F) The expected external disturbances (sizing).

All of these factors are inter-related and the tradeoff of one factor against another depends upon the weighting that the particular application demands. For example, the configurations that exhibit good momentum utilization (i.e. the configuration which will allow the total momentum to be directed completely along any axes) generally requires a more complicated control law to reduce cross coupling.

Figures 2.2 through 2.6 illustrate a few of the basic configurations that can be used for three axis control and Table 2-1 shows a comparison of these configurations relative to the system considerations. There are many other configurations that can be used depending on the particular application and there is no intent that the configurations shown define all possible configurations.(2)

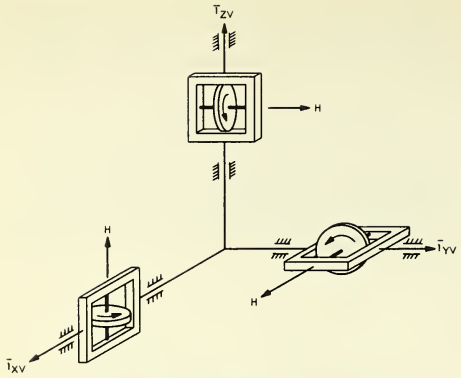


FIGURE 2.2 THREE SINGLE GIMBAL CMG

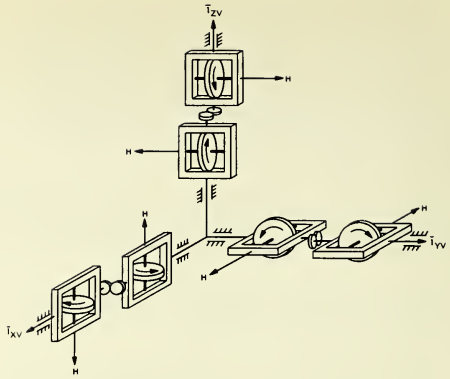


FIGURE 2.3 THREE SINGLE GIMBAL CMG PAIRS

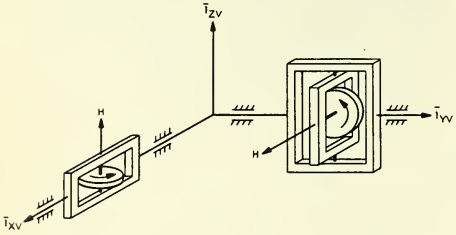


FIGURE 2.4 ONE DOUBLE GIMBAL CMG AND ONE SINGLE GIMBAL CMG

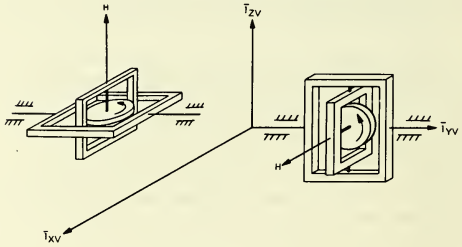


FIGURE 2.5 TWO DOUBLE GIMBAL CMGs

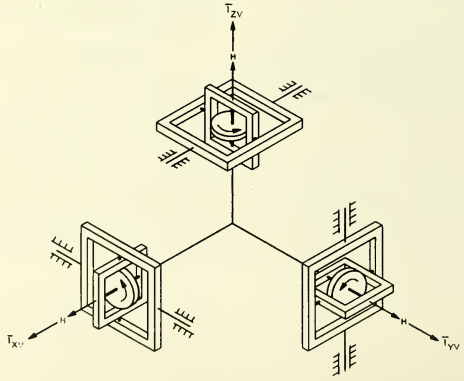


FIGURE 2.6 THREE DOUBLE GIMBAL CMG (SIX-PAC) CONFIGURATION

CONFIGURATION**	REFERENCE FIGURE	SIZE AND WEIGHT*	POWER†	MOMENTUM UTILIZATION (% OF TOTAL MOMENTUM)	REDUNDANCY (% OF TOTAL CONTROL)
3 SCMG's	2-2	2	1.5	67 each axis	0
3 Pairs of SCMG's	2-3	5	3	33 each axis	100
1 CMG and 1 SCMG	2-4	1	1	100 for 2 axes 50 for 1 axis	0
2 CMG	2-5	3	1	100 for all axes	25
3 CMG (SIX-PAC)	2-6	4	1.5	100 for all axes	100

*The scale is for comparison in proportion to actual weight of 1 CMG and 1 SCMG
**SCMG means single gimbal CMG
CMG means double gimbal CMG
†This scale is for comparison in proportion to actual power of 1 CMG and 1 SCMG

TABLE 2-1. COMPARISON OF CMG CONFIGURATIONS AGAINST SYSTEM CONSTRAINTS FOR THREE AXIS CONTROL

The function of the CMG control law as stated previously is to convert the command moment into the proper CMG control (command) variable such that the desired reaction moment will be obtained by the vehicle. Hence, the nature of the control law depends on the mode of control and the CMG dynamics. There are three basic CMG modes of control which are named according to the type of control variable:

A) Torque

B) Rate

C) Position.

The torque mode of operation has a moment applied to the CMG pivot by a torquer which is balanced by a reaction torque caused by the gyroscopic action. This mode is open loop and degenerates if there is appreciable friction on the gimbal pivots and thus is basically not usable.

The rate mode of operation can be implemented in either a closed or open loop configuration and the controlled variable is the time rate of change of the total CMG angular momentum vector $\bar{H}_T$. The required $\bar{H}_T$ vector is obtained by controlling the CMG gimbal rates as a function of the input variable $\bar{H}_{T(com)}$.

The position mode is a closed loop controller and the controlled variable is the total CMG angular momentum vector $\bar{H}_T$. The required $\bar{H}_T$ vector is obtained by controlling the CMG gimbal angles as a function of the input variable $\bar{H}_{T(com)}$.

The perfect control law produces a moment from the CMG, $\bar{M}_{R(CMG)}$, equal to the command moment $\bar{M}_{(com)}$ or

$$\bar{M}_{(com)} = \bar{M}_{R(CMG)} \tag{2.2}$$

A closed loop rate controller will ideally implement equation (2.3).

$$\bar{M}_{(com)} = \left\{ \frac{d\,\bar{H}_{T(com)}}{dt} \right\}_I = -\left\{ \frac{d\,\bar{H}_{T(CMG)}}{dt} \right\}_I$$

$$= \bar{M}_{R(CMG)} \tag{2.3}$$

A closed loop position controller will ideally implement equation (2.4) which is obtained by integrating equation (2.3),

$$\bar{H}_{T(com)} = \bar{H}_{T(CMG)} + \bar{H}_{T(CMG)}^{(0+)}, \tag{2.4}$$

where $\bar{H}_{T(CMG)}^{(0+)}$, is the initial momentum of the CMG configuration.

The closed loop rate and position controllers will approximate the ideal control law as defined by equation (2.2). An open loop rate mode in general is simpler, however, it will produce a

control law with large gain variation and considerable cross coupling moments.

The actual forms of the control laws or combinations of control laws will depend on the particular CMG cluster configurations. A particular application is the Apollo Telescope Mount (ATM) System which will be discussed in more detail in subsequent sections.

Momentum Management

Momentum management is the method of maintaining the total momentum of the momentum controller within its momentum volume and is subdivided into the following three functions:

A) Saturation measurement

B) Desaturation computation

C) Desaturation implementation.

Saturation measurement involves the measurement of the controller momentum components along a given reference coordinate system (usually the vehicle principle control axes) and combining these components so as to determine and display the state of the momentum controller relative to the controller momentum volume. The complexity of this function varies widely, from simple limit switches to full momentum volume indication.

Desaturation computation contains the logic and computations necessary to affect the desired momentum state to which the controller is driven and the determination of the initial and final desaturation commands. The computations of the desired momentum state can be as simple as commanding the total momentum to zero (likely for random disturbances) or as sophisticated as optimizing the momentum control capability by accounting for past momentum profiles.

Desaturation implementation is the actual mechanization of forcing the momentum controller to attain the momentum state determined by the desaturation computations.

For simple desaturation schemes this function is accomplished by electrically caging each CMG gimbal angle in the controller cluster to prescribed gimbal angles. More sophisticated desaturation implementation utilizes the control loop functions to drive the CMGs, in concert, to attain the desired momentum in a momentum feedback loop. The latter implementation scheme is used in the ATM fine control system.

2.3 THE ATM ATTITUDE CONTROL SYSTEM

The ATM Program[3] is an example of the use of CMGs for the control of a large spacecraft. Figure 2.7 depicts one of the ATM vehicle configurations to be controlled by a cluster of CMGs.

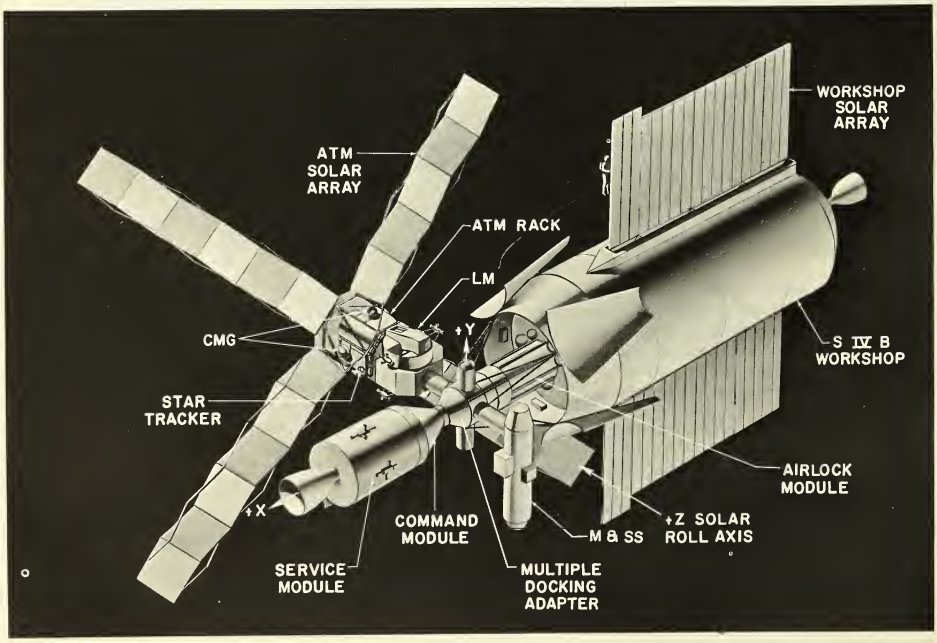


FIGURE 2.7 APOLLO EXPERIMENT STATION
Courtesy of Marshall Space Flight Center

The ATM pointing control system will utilize a cluster of CMGs located on the ATM rack as shown in Figure 2.7.

The CMG cluster for the ATM pointing control system is the SIX-PAC[1] configuration (also referred to as the Langley Configuration) and is shown schematically in Figure 2.8. It consists of three two-degree of freedom CMGs which are oriented with respect to the vehicle coordinates as defined by the transformation matrices given in Equations (2.5) through (2.7).

$$\begin{pmatrix} H_{XV(1)} \\ H_{YV(1)} \\ H_{ZV(1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} X \begin{pmatrix} H_{1B(1)} \\ H_{2B(1)} \\ H_{3B(1)} \end{pmatrix} \quad (1)\ \text{CMG}\ (2.5)$$

$$\begin{pmatrix} H_{XV(2)} \\ H_{YV(2)} \\ H_{ZV(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} X \begin{pmatrix} H_{1B(2)} \\ H_{2B(2)} \\ H_{3B(2)} \end{pmatrix} \quad (2)\ \text{CMG}\ (2.6)$$

$$\begin{pmatrix} H_{XV(3)} \\ H_{YV(3)} \\ H_{ZV(3)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} X \begin{pmatrix} H_{1B(3)} \\ H_{2B(3)} \\ H_{3B(3)} \end{pmatrix} \quad (3)\ \text{CMG}\ (2.7)$$

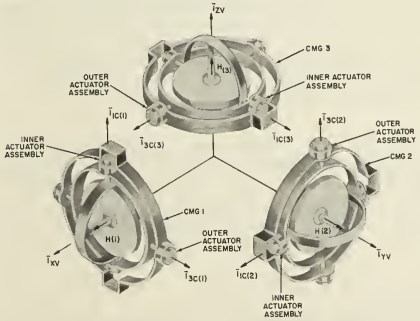


FIGURE 2.8 ATM CMG (SIX-PAC) CONFIGURATION

This configuration provides full momentum utilization along any axis and readily allows redundant operation which consists of any two CMGs in the active state and the third CMG not operating.

The ATM fine point system has the capability of implementing either the momentum position control, $\bar{H}$-vector, or the $\dot{\bar{H}}$ vector control laws. The momentum position control, which was conceptually described in paragraph 2.2, offers the least cross coupling and the least complexity, and it can be used directly for desaturation with minimal mode switching.

A block diagram of the $\bar{H}$-vector position control law for the ATM fine pointing system is shown in Figure 2.9. The moment command vector, $\bar{M}_{(com)}$, determined by vehicle control law is integrated to yield a command controller momentum, $\bar{H}_{T(com)}$. The initial conditions on this integration are determined by the momentum management control. This command momentum is compared with the actual momentum, $\bar{H}_T$, and the difference or error momentum, $\bar{H}_{TE}$, is used as a modified command in a cross product CMG steering law which is defined in equation (2.8), where $\bar{W}_{VA(j)}$ is the j^{th} CMG inner gimbal angular rate

$$\bar{W}_{VA(j)} = \left\{ \bar{I}_{2A(j)} \times \bar{H}_{TE} \right\} \qquad j = 1, 2, 3. \qquad (2.8)$$

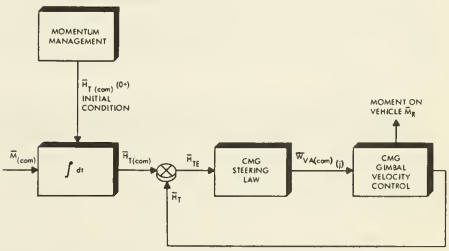


FIGURE 2.9 H-VECTOR CONTROL LAW

The CMG cross product steering law generates six individual CMG rate commands which are used to move the CMG momentum vectors and thus produce a moment on the vehicle. The new gimbal positions of the CMGs are used to measure the controller total momentum vector $\bar{H}_T$ which closes the loop.

The momentum position control law approximates the ideal control law and is discussed in detail in section 3.

The $\dot{\bar{H}}$-vector closed loop control law is quite similar to the momentum vector position control law and approximates the ideal control law; however, its mechanization is more complex. The $\dot{\bar{H}}$-vector open loop control law provides the simplest mechanization, but is characterized by gain variations and cross coupling moments and thus does not approach the ideal characteristics that are exhibited by either closed loop control law.

The $\dot{\bar{H}}$-vector control is discussed in detail in section 4. and the ATM Attitude Control System is

capable of implementing any one of the three types of control laws.

The CMG configuration shown in Figure 2.8 allows much versatility in implementation of momentum management because of the 100% momentum utilization, and hence the spherical momentum volume.

Saturation measurement consists of displaying the components of $\bar{H}_T$ and the computation of momentum magnitude $|\bar{H}_T|$ for the use as a gas gage indication, where $|\bar{H}_T|$ is defined as

$$|\bar{H}_T| = \left[H_{TXV}^2 + H_{TYV}^2 + H_{TZV}^2\right]^{1/2} \qquad (2.9)$$

and H_{TXV}, H_{TYV}, H_{TZV} are defined as the X, Y, Z components of the $\bar{H}_T$ vector in vehicle space.

Desaturation computation provides saturation indication and initialization on command when $|\bar{H}_T| \geq \epsilon$, where ϵ is a preset value.

The value of ϵ is largely a function of the maximum moment change during one orbit due to an apriority evaluation of external disturbances. This value can be updated as a result of actual measurements of $|\bar{H}_T|$.

If $|\bar{H}_T|$ is less than ϵ, the astronaut knows that he can operate for at least one more orbit, if $|\bar{H}_T|$ is greater than or equal to ϵ, he knows that he must command desaturation at the first operational opportunity within his present orbit.

The desaturation computation function determines the desired value of $\bar{H}_T$ at the completion of desaturation. This value of $\bar{H}_T$ is denoted as $\bar{H}_{T(com)}^{(0+)}$.

The implementation of desaturation is accomplished by forcing the output of the integrator (Figure 2-9), $\bar{H}_{T(com)}$, to

$$\bar{H}_{T(com)} = \bar{H}_{T(com)}^{(0+)} . \qquad (2.10)$$

Note: This implementation is the initial condition in equation (2.4).

Desaturation is deactivated when

$$\bar{H}_{TE} = \bar{H}_{T(com)}^{(0+)} - \bar{H}_T \leq \bar{\eta} \qquad (2.11)$$

where $\bar{\eta}$ is an acceptable error vector and is not a critical parameter.

3. AN ANGULAR MOMENTUM VECTOR POSITION CONTROLLER

The reaction moment on a spacecraft from an ideal cluster (as shown in Figure 2.8) of CMGs can be approximated as the time derivative of the total angular momentum vector with respect to vehicle space or

$$\bar{M}_R \approx - \left\{\frac{d\,H_T}{d\,t_V}\right\} . \qquad (3.1)$$

From equation (3.1) the total angular momentum vector $\bar{H}_T$ of a cluster of CMGs is equal to the time integral of the reaction moment exerted on the vehicle or

$$\bar{H}_T = - \int_{t_1}^{t_2} \bar{M}_R \, dt . \qquad (3.2)$$

Assume that a controller can be implemented so that the desired total angular momentum vector, $\bar{H}_{T(com)}$, is equal to the actual total angular momentum vector, $\bar{H}_T$, then reaction moment on the spacecraft can be properly controlled as a function of the attitude control system error signal. A simplified block diagram of a total angular momentum controller is shown in Figure 3.1.

The $\dot{\bar{H}}_{T(com)}$ vector is the rate of change of $\bar{H}_{T(com)}$ with respect to vehicle or "V" space and is proportional to a linear combination of attitude error and vehicle body axis rate. Each component of this vector is separately integrated to form

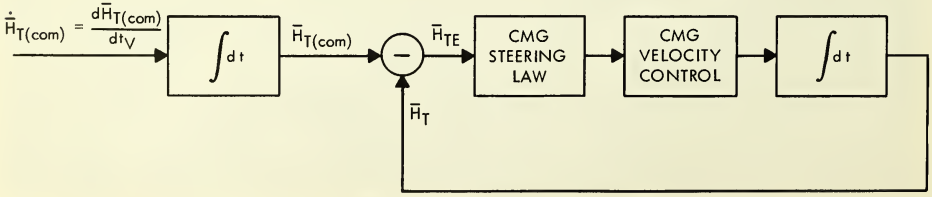


FIGURE 3.1 THE $\bar{H}_T$ CONTROLLER BLOCK DIAGRAM

the $\bar{H}_{T(com)}$ vector. This signal is compared in vector form to the measured total momentum vector, $\bar{H}_T$, to obtain an error signal $\bar{H}_{TE}$. The error signal will be processed by a steering law which is a vector cross product $\left\{\bar{I}_{H(1)} \times \bar{H}_{TE}\right\}$, $\left\{\bar{I}_{H(2)} \times \bar{H}_{TE}\right\}$ and $\left\{\bar{I}_{H(3)} \times \bar{H}_{TE}\right\}$ for the (1), (2), and (3) CMGs respectively. The output of the steering law will drive the CMG gimbal velocity servos and thus position the angular momentum vector of each CMG to form the vector $\bar{H}_T$.

3.1 THE EQUATIONS OF THE ANGULAR MOMENTUM POSITION CONTROLLER

The reaction moment of an ideal cluster of CMGs on the vehicle is expressed in equation (3.3),

$$\bar{M}_R = - \left\{\frac{d\,\bar{H}_T}{d\,t_V} + \bar{W}_{IV} \times \bar{H}_T\right\} \qquad (3.3)$$

where an ideal CMG is defined as one that has instantaneous servo loops and no inertia reaction moments. The vector $\bar{H}_T$ is the total angular momentum of the three CMGs as shown in equation (3.4).

$$\bar{H}_T = \sum_{j=1}^{3} \bar{H}_{(j)} \qquad (3.4)$$

If one assumes that the vehicle rates are small, then the cross product term $\left\{\bar{W}_{IV} \times \bar{H}_T\right\}$ in equation (3.3) can be neglected and the reaction moment is expressed as equation (3.5).

$$\bar{M}_R = - \frac{d\,\bar{H}_T}{d\,t_V} = - \sum_{j=1}^{3} \frac{d\,\bar{H}_{(j)}}{d\,t_V} \qquad (3.5)$$

Equation (3.5) can be expanded as shown in equation (3.6),

$$\bar{M}_R = - \sum_{j=1}^{3} \left\{\frac{d\,\bar{H}_{(j)}}{d\,t_{A(j)}} + \bar{W}_{VA(j)} \times \bar{H}_{(j)}\right\} \qquad (3.6)$$

where "A(j)" space is inner gimbal space of the (j) CMG. Since the angular momentum vector of a CMG is a constant in "A" space the time derivatives in equation (3.6) are zero and thus equation (3.6) can be rewritten as equation (3.7).

$$\bar{M}_R = - \frac{d\,\bar{H}_T}{d\,t_V} = - \sum_{j=1}^{3} \bar{W}_{VA(j)} \times \bar{H}_{(j)} \qquad (3.7)$$

Let the steering law for the j^{th} CMG be defined as shown in equation (3.8),

$$\bar{W}_{VA(j)} = K_{s1} \left\{\bar{I}_{H(j)} \times \bar{H}_{TE}\right\} \qquad (j)\ \text{CMG} \qquad (3.8)$$

where j = 1, 2, 3 and the vector $\bar{H}_{TE}$ is defined in equation (3.9).

$$\bar{H}_{TE} = \bar{H}_{T(com)} - \bar{H}_T \qquad (3.9)$$

Substituting equation (3.8) in (3.7) and expanding the triple cross product terms, the vector $\left\{d\,\bar{H}_T/d\,t_V\right\}$ can be expressed as

$$\frac{d\,\bar{H}_T}{d\,t_V} = H\,K_{s1} \left\{3\,\bar{H}_{TE} - \sum_{j=1}^{3} \left[\bar{I}_{H(j)} \cdot \bar{H}_{TE}\right] \bar{I}_{H(j)}\right\}, \qquad (3.10)$$

where H is defined as the magnitude of the angular momentum vectors of the (1), (2) and (3) CMGs and

$$H = \left|\bar{H}_{(j)}\right|, \quad j = 1, 2, 3. \qquad (3.11)$$

An ideal first order equation to control the vector $\bar{H}_T$ can be written as shown in equation (3.12).

$$\frac{d\,\bar{H}_T}{d\,t_V} = K\,\bar{H}_{TE} \qquad (3.12)$$

The controller described in equation (3.12) is stable and the vector $\bar{H}_{TE}$ for a constant command moment is zero in the steady state. If K is sufficiently large, then $\bar{H}_{T(com)}$ will be equal to $\bar{H}_T$ for all desired frequencies and the reaction moment, $\bar{M}_R$, is equal to $\dot{\bar{H}}_{T(com)}$ as shown in equation (3.13).

$$\dot{\bar{H}}_{T(com)} = \frac{d\,\bar{H}_T}{d\,t_V} = - \bar{M}_R \qquad (3.13)$$

Equations (3.10) and (3.12) are quite similar and the steady state value of $\bar{H}_{TE}$ for equation (3.10) will be zero for all cases except when all three vectors $\bar{H}_{(1)}$, $\bar{H}_{(2)}$ and $\bar{H}_{(3)}$ are colinear with the error vector $\bar{H}_{TE}$. When this colinear condition is obtained $\left\{d\,\bar{H}_t/d\,t_V\right\}$ is zero as can be seen by equations (3.7) and (3.8), thus it is impossible to correct the error vector $\bar{H}_{TE}$.

There are two cases of colinearity, one when all three CMG angular momentum vectors are aligned in the same direction and the other when two CMG angular momentum vectors are aligned in the same

direction and the third CMG has its angular momentum vector opposing the other two. These two cases are shown in Figures 3.2A and 3.2B:

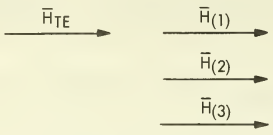


FIGURE 3.2A THE 3H CONFIGURATION

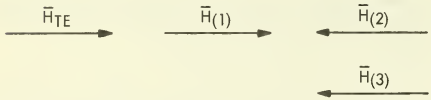


FIGURE 3.2B A COLINEAR 1H CONFIGURATION

The 3 H configuration shown in Figure 3.2A shows the CMG system at its maximum capacity and it must be desaturated by utilization of controlled external moments such as reaction control systems or gravity gradients. The 1 H configuration, shown in figure 3-2B, is an unstable configuration and the system noise will cause the $\bar{H}_{(2)}$ and $\bar{H}_{(3)}$ vectors to reposition themselves so that they are no longer colinear with the original value of $\bar{H}_{TE}$, and thus the final value of $\bar{H}_{TE}$ will be zero.

3.2 IMPLEMENTATION OF THE $\bar{H}_T$ VECTOR CONTROL LAW

The complete controller for the ATM/CMG cluster is shown in Figure 3.3. The vector $\bar{H}_T$ is the sum of the $\bar{H}_{(1)}$, $\bar{H}_{(2)}$, and $\bar{H}_{(3)}$ angular momentum vectors as shown in equation (3.4), and this summation will be made in "V" or vehicle space. The vectors $\bar{H}_{(1)}$, $\bar{H}_{(2)}$, and $\bar{H}_{(3)}$ are first expressed in inner gimbal space of their respective individual CMGs as shown in equation (3.14).

$$\bar{H}_{(j)} = H \left\{ \bar{I}_{2A(j)} \right\}, \quad j = 1, 2, 3. \qquad (3.14)$$

Each vector is transformed from inner gimbal space to CMG base space by means of a resolver chain mounted on the CMG pivots, and is shown by the lower resolver chain in Figure 3.3.

The components of $\bar{H}_{(1)}$ and $\bar{H}_{(3)}$ in CMG base space are suitably connected to the X, Y, and Z summing amplifiers to account for their base orientation with respect to the vehicle coordinate system. This method of interconnection can be obtained from the transformation matrices given in equations (2.5), (2.6), and (2.7).

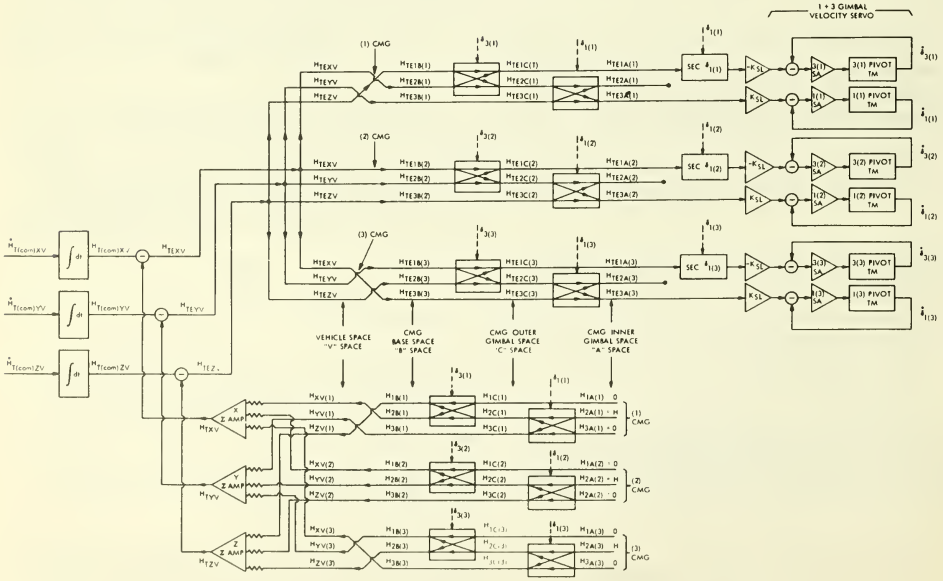


FIGURE 3.3 A BLOCK DIAGRAM OF AN ANGULAR MOMENTUM VECTOR POSITION CONTROLLER

The output of the X, Y, and Z summing amplifiers are then compared to the input command signals to obtain the three error signals H_{TEXV}, H_{TEYV} and H_{TEZV} as shown in Figure 3.3.

The output signals from the steering law are best implemented by expressing the vector $\bar{H}_{TE}$ in inner gimbal space for each CMG. The vector $\bar{H}_{TE}$ is transformed from vehicle to CMG base space as defined by equations (2.5), (2.6), and (2.7) to account for the (1) and (3) CMG orientation with respect to the vehicle coordinates. The error vector in CMG base coordinates is then transformed to CMG inner gimbal or "A" space by means of a resolver chain, as shown by the upper resolver chain of Figure 3.3, so that $\bar{H}_{TE}$ is expressed as

$$\bar{H}_{TE} = \sum_{i=1}^{3} \left\{ \bar{l}_{iA(j)} \; H_{TEiA(j)} \right\} , \quad j = 1, 2, 3. \tag{3.15}$$

The steering law for the j^{th} CMG is given in equation (3.8), and by substituting equation (3.15) in equation (3.8) the required inner gimbal rate of the j^{th} CMG can be expressed as equation (3.16).

$$\bar{W}_{VA(com)(j)} = K_{s1} \left[\bar{l}_{1A(j)} \; H_{TE3A(j)} - \bar{l}_{3A(j)} \; H_{TE1A(j)} \right] \tag{3.16}$$

The terms $H_{TE3A(j)}$ and $H_{TE1A(j)}$ are the input command signals to the (1) and (3) pivots of the j^{th} CMG.

The angular velocity of the inner gimbal of the j^{th} CMG can be expressed as a function of relative gimbal angle rates as shown in equation (3.17).

$$\bar{W}_{VA(j)} = \bar{l}_{1A(j)} \; \dot{\delta}_{1(j)} + \bar{l}_{2A(j)} \left[\sin \delta_{1(j)} \right] \dot{\delta}_{3(j)} + \bar{l}_{3A(j)} \left[\cos \delta_{1(j)} \right] \dot{\delta}_{3(j)} . \tag{3.17}$$

Assume that the j^{th} CMG has two gimbal velocity servos that are infinitely fast and described by equation (3.18).

$$K_{s1} \; H_{TE3A(j)} = \dot{\delta}_{1(j)}$$

$$-K_{s1} \sec \delta_{1(j)} \; H_{TE1A(j)} = \dot{\delta}_{3(j)} \tag{3.18}$$

Substituting equation (3.18) into (3.17) the inner gimbal rate of the j^{th} CMG can be expressed as equation (3.19).

$$\bar{W}_{VA(j)} = K_{s1} \left\{ \bar{l}_{1A(j)} \; H_{TE3A(j)} - \bar{l}_{2A(j)} \left[\tan \delta_{1(j)} \; H_{TE1A(j)} \right] - \bar{l}_{3A(j)} \; H_{TE1A(j)} \right\} . \tag{3.19}$$

The vectors $\bar{W}_{VA(j)(com)}$ and $\bar{W}_{VA(j)}$ are not identical as can be seen from equations (3.16) and (3.19). The actual rate of the inner gimbal has a component along the $\bar{l}_{2A(j)}$ vector while the command rate of the inner gimbal has components only along the $\bar{l}_{1A(j)}$ and $\bar{l}_{3A(j)}$ vectors. The desired reaction moment from $\bar{W}_{VA(com)(j)}$ is obtained by substituting equation (3.16) into (3.7) and the actual reaction moment from $\bar{W}_{VA(j)}$ is obtained by substituting equation (3.19) into (3.7). The result of either substitution in equation (3.7) is given in equation (3.20),

$$\frac{d \bar{H}_T}{d t_V} = K_{s1} \sum_{j=1}^{3} \left\{ \bar{l}_{1A(j)} \; H_{TE1A(j)} + \bar{l}_{3A(j)} \; H_{TE3A(j)} \right\} \tag{3.20}$$

thus the extra component in $\bar{W}_{VA(j)}$ along the $\bar{l}_{2A(j)}$ vector has no effect on the controller.

The gimbal velocity servo equations (equations (3.10) through (3.18) are implemented in the top half of Figure 3.3, however; it can be shown that the secant term can be approximated by a constant gain since the inner pivot angle has a limited range of ±70 degrees.

The saturation measurement of the system momentum can be obtained by squaring the outputs of the X, Y, and Z summing amplifiers as shown in Figure 3.3 and adding three resulting signals.

The $\bar{H}_T$ controller will operate in a backup mode with any two CMGs operating and there is no major switching required between a three CMG and a two CMG cluster mode of operation. As two units are operating, the open loop gain of the controller is reduced to two-thirds of the gain with three CMGs in operation; however, the total closed loop gain remains the same for either two or three CMG operation.

4. AN ANGULAR MOMENTUM VECTOR RATE CONTROL

The reaction moment on a spacecraft can also be controlled in an open loop mode as a function of the $\bar{H}_{T(com)}$ vector. The $\dot{\bar{H}}_{T(com)}$ vector, which is proportional to a linear combination of attitude error and vehicle body axis rate, is the input signal to a steering law mechanism as shown in Figure 4.1.

This input signal contains three variables which must be suitably processed to provide six output signals for the six gimbal velocity servos of the

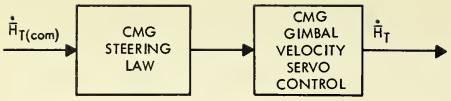


FIGURE 4.1 AN OPEN LOOP $\dot{\bar{H}}_T$ CONTROL

(1), (2) and (3) CMGs. The output of the CMG velocity servos provides the rate of change of $\bar{H}_T$ or $\{d\bar{H}_T/dt_V\}$ and the desired value of $\bar{H}_T$ will then be obtained through the attitude control system by a change in attitude of the spacecraft.

4.1 THE EQUATIONS OF THE OPEN LOOP ANGULAR MOMENTUM VECTOR CONTROL

The steering law for the j^{th} CMG is similar in form to that used in the $\bar{H}_T$ controller and is described in equation (4.1).

$$\bar{W}_{VA(j)} = K_{s1} \left\{\bar{I}_{H(j)} \times \dot{\bar{H}}_{T(com)}\right\}, \quad (j^{th}) \text{ CMG} \tag{4.1}$$

However, the input variable is $\dot{\bar{H}}_{T(com)}$ rather than $\bar{H}_{TE}$ as shown in equation (3.8). Substituting equation (4.1) in (3.7) $\{d\bar{H}_T/dt_V\}$ can be expressed as equation (4.2),

$$\frac{d\bar{H}_T}{dt_V} = H K_{s1} \left\{3\dot{\bar{H}}_{T(com)} - \sum_{j=1}^{3} \left[\bar{I}_{H(j)} \cdot \dot{\bar{H}}_{T(com)}\right] \bar{I}_{H(j)}\right\} \tag{4.2}$$

where H is defined as the magnitude of the angular momentum of each CMG. If the gain K_{s1} is equal to one-third H then equation (4.2) can be written as

$$\frac{d\bar{H}_T}{dt_V} = \dot{\bar{H}}_{T(com)} - \frac{1}{3}\sum_{j=1}^{3} \left[\bar{I}_{H(j)} \cdot \dot{\bar{H}}_{T(com)}\right] \bar{I}_{H(j)} \tag{4.3}$$

As can be seen from equation (4.3) if $\dot{\bar{H}}_{T(com)}$ were perpendicular to the three angular momentum vectors $\bar{H}_T$ would equal $\dot{\bar{H}}_{T(com)}$ which is the desired relationship. As the orientation of $\dot{\bar{H}}_{T(com)}$ varies with respect to the angular momentum vectors both gain variations and cross coupling can be expected.

4.2 THE $\dot{\bar{H}}_T$ LOOP IMPLEMENTATION

The implementation of the $\dot{\bar{H}}_T$ open loop controller is shown in Figure 4.2 and is quite similar to the implementation shown in Figure 3.3. The steering law is implemented in the top half of the diagram by transforming the $\dot{\bar{H}}_{T(com)}$ vector from vehicle space to inner gimbal space of each CMG. The gimbal velocity servos are described by equation (4.4) and are similar to those used in the H_T controller of Figure 3.3.

$$K_{s1}\, \dot{H}_{T(com)3A(j)} = \dot{\delta}_{1(j)}$$

$$-K_{s1} \sec \delta_{1(j)}\, \dot{H}_{T(com)1A(j)} = \dot{\delta}_{3(j)} \tag{4.4}$$

Equation (4.5) describes the gimbal relative rates as a function of $\dot{\bar{H}}_{T(com)}$ and defines the steering law implemented in Figure 4.2.

$$\begin{bmatrix} \dot{\delta}_{1(1)} \\ \dot{\delta}_{1(2)} \\ \dot{\delta}_{1(3)} \\ \dot{\delta}_{3(1)} \\ \dot{\delta}_{3(2)} \\ \dot{\delta}_{3(3)} \end{bmatrix} = \frac{1}{3H} \begin{bmatrix} -\{\cos\delta_{3(1)} \sin\delta_{1(1)}\} & \cos\delta_{1(1)} & \{\sin\delta_{3(1)} \sin\delta_{1(1)}\} \\ \{\sin\delta_{3(2)} \sin\delta_{1(2)}\} & -\{\cos\delta_{3(2)} \sin\delta_{1(2)}\} & \cos\delta_{1(2)} \\ \cos\delta_{1(3)} & \{\sin\delta_{3(3)} \sin\delta_{1(3)}\} & -\{\cos\delta_{3(3)} \sin\delta_{1(3)}\} \\ -\{\sin\delta_{3(1)} \sec\delta_{1(1)}\} & 0 & -\{\cos\delta_{3(1)} \sec\delta_{1(1)}\} \\ -\{\cos\delta_{3(2)} \sec\delta_{1(2)}\} & -\{\sin\delta_{3(2)} \sec\delta_{1(2)}\} & 0 \\ 0 & -\{\cos\delta_{3(3)} \sec\delta_{1(3)}\} & -\{\sin\delta_{3(3)} \sec\delta_{1(3)}\} \end{bmatrix} \begin{bmatrix} \dot{H}_{T(com)XV} \\ \dot{H}_{T(com)YV} \\ \dot{H}_{T(com)ZV} \end{bmatrix} \tag{4.5}$$

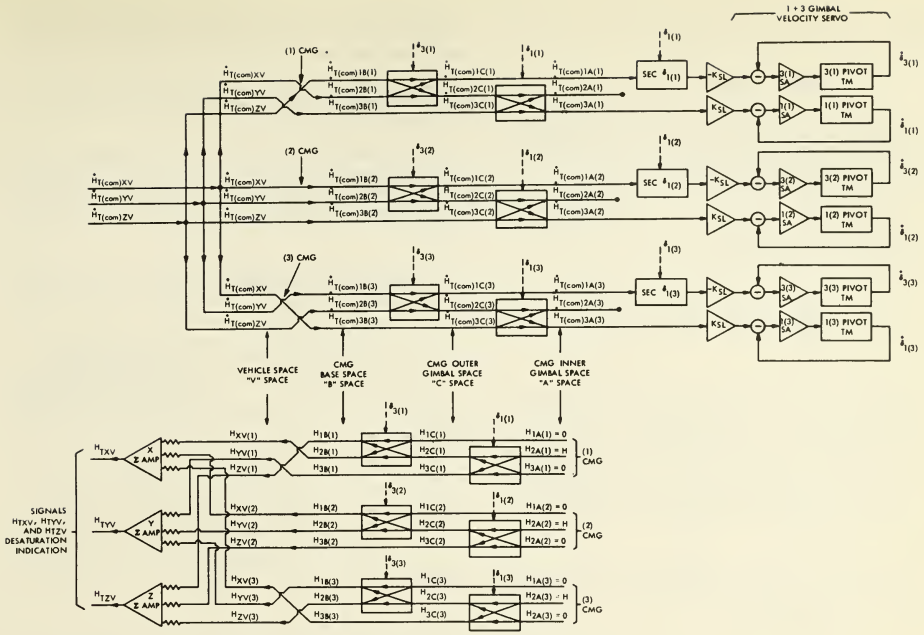


FIGURE 4.2 A BLOCK DIAGRAM OF AN OPEN LOOP $\dot{\bar{H}}_T$ CONTROLLER

Equation (4.5) or the implementation shown in Figure 4.2 is an exact expression of a steering law defined in a NASA Langley report.[1]

The lower half of Figure 4.2 defines a system for measuring H_{TXV}, H_{TYV} and H_{TZV} and is the same as described in section 3.2. The three components are squared and summed to define the saturated 3 H configuration of $\bar{H}_T$.

Again, this system can operate in a secondary mode with any two CMGs operating without any major switching requirements.

4.3 THE CLOSED LOOP $\dot{\bar{H}}_T$ VECTOR CONTROLLER

The gain variations and cross coupling of the $\dot{\bar{H}}_T$ open loop controller (see equation (4.3)) can be minimized by the use of a closed loop system as shown in Figure 4.3.

This configuration will approximate many of the characteristics of an ideal control law and is similar to the $\bar{H}_T$ controller. The steering law for the j^{th} CMG is similar to equation (4.1) and can be expressed as

$$\bar{W}_{VA(j)} = K_{sl} \left\{ \bar{I}_{H(j)} \times \dot{\bar{H}}_{TE} \right\} \quad (j^{th} \text{ CMG}) \qquad (4.6)$$

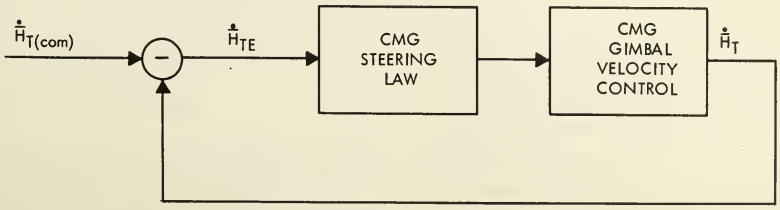


FIGURE 4.3 A CLOSED LOOP $\dot{\bar{H}}_T$ CONTROLLER

67-589

where the error vector $\bar{\dot{H}}_{TE}$ is defined as equation (4.7).

$$\bar{\dot{H}}_{TE} = \bar{\dot{H}}_{T(com)} - \bar{\dot{H}}_{T} \tag{4.7}$$

Substituting equation (4.6) in (3.7), the differential equation of the closed loop controller can be expressed as equation (4.8).

$$\frac{d\,\bar{H}_T}{d\,t_V} = H K_{s1} \left\{ 3\bar{\dot{H}}_{TE} - \sum_{j=1}^{3} \left[\bar{l}_{H(j)} \cdot \bar{\dot{H}}_{TE} \right] \bar{l}_{H(j)} \right\} \tag{4.8}$$

If K_{s1} is equal to 10/H, then the closed loop gain will vary from 0.97 when all momentum vectors are perpendicular to the $\bar{\dot{H}}_{TE}$ vector, to 0.64 when all momentum vectors are 20^{o} from the $\bar{H}_{TE}$ vector. The closed loop variation of 1.52:1 can be compared to an open loop gain variation of 16.6:1, thus the closed loop control of $\bar{\dot{H}}_T$ will effectively stabilize the open loop gain.

The gimbal velocity servos are similar to those described in equations (3.18) and (4.4), and are described in equation (4.9).

$$K_{s1}\, \dot{H}_{TE3A(j)} = \dot{\delta}_{1(j)}$$

$$-K_{s1} \sec \delta_{1(j)}\, \dot{H}_{TE1A(j)} = \dot{\delta}_{3(j)} \tag{4.9}$$

The measured value of $\bar{\dot{H}}_T$ is obtained by substituting equation (3.17) in equation (3.7) and it can be expressed as equation (4.10).

$$\bar{\dot{H}}_T = H \sum_{j=1}^{3} \left\{ - \bar{l}_{1A(j)} \left[\cos \delta_{1(j)} \right] \dot{\delta}_{3(j)} + \bar{l}_{3A(j)}\, \dot{\delta}_{1(j)} \right\} \tag{4.10}$$

Figure 4.4 is an implementation of the closed loop H_T controller, which is described by equations (4.6), (4.7), (4.9) and (4.10). The top resolver chain in Figure 4.4 mechanizes equations (4.6), (4.7) and (4.9) in a similar manner to that described for the $\bar{H}_T$ controller which is shown in Figure 3.3. Each vector $\left\{ - \bar{l}_{1A(j)} \left[\cos \delta_{1(j)} \right] \dot{\delta}_{3(j)} + \bar{l}_{3A(j)}\, \dot{\delta}_{1(j)} \right\}$ of equation (4.10) is transformed from CMG inner gimbal space to vehicle space by means of the middle resolver chain of Figure 4.4. The individual components of the vectors are appropriately summed by summing amplifiers and each amplifier has a gain of H.

The lower resolver chain in Figure 4.4 instruments the measurement of $\bar{H}_T$ which is used for momentum management.

The closed loop $\bar{\dot{H}}_T$ controller is somewhat more complex than the $\bar{H}_T$ controller since it requires an additional resolver chain as shown in Figure 4.4 which replaces the three integrators shown in Figure 3.3.

5. THE KINEMATIC EQUATIONS OF MOTION OF A CMG

The exact equations of motion of a CMG are quite complex and difficult to use in an attitude control system simulation. If certain simplifying assumptions are made, a reasonable mathematical model can be developed for simulation and stability studies.

The CMG is a two degree of freedom gyro and is shown schematically in Figure 1.3. Its inner gimbal is defined as "A" space, the outer gimbal as "C" space and the base as "B" space. The inner and outer pivots are designated as the (1) and (3) pivots, and each pivot is driven by a geared torque motor.

The derivation can be divided into four parts:

A) The equations of motion of the inner gimbal written in "C" space

B) The geared torque motor equations

C) The inner and outer gimbal servo loop kinematic equations

D) The CMG reaction moment exerted on the base written in "C" space.

The following definitions of polar moments of inertia will be used throughout this analysis.

$J_R \triangleq$ The polar moment of inertia of the wheel about the $\bar{l}_{2A}$ vector.

$J_D \triangleq$ The polar moment of inertia of the wheel about an axis perpendicular to the $\bar{l}_{2A}$ vector.

$J_{A11}, J_{A22}, J_{A33} \triangleq$ The polar moments of inertia of the inner gimbal exclusive of the wheel about the $\bar{l}_{1A}$, $\bar{l}_{2A}$, $\bar{l}_{3A}$ vectors respectively.

$J_{C33} \triangleq$ The polar moment of inertia of the outer gimbal about the $\bar{l}_{3C}$ vector.

$J_{MR} \triangleq$ The polar moment of inertia of the torque motor rotor about its rotational axis.

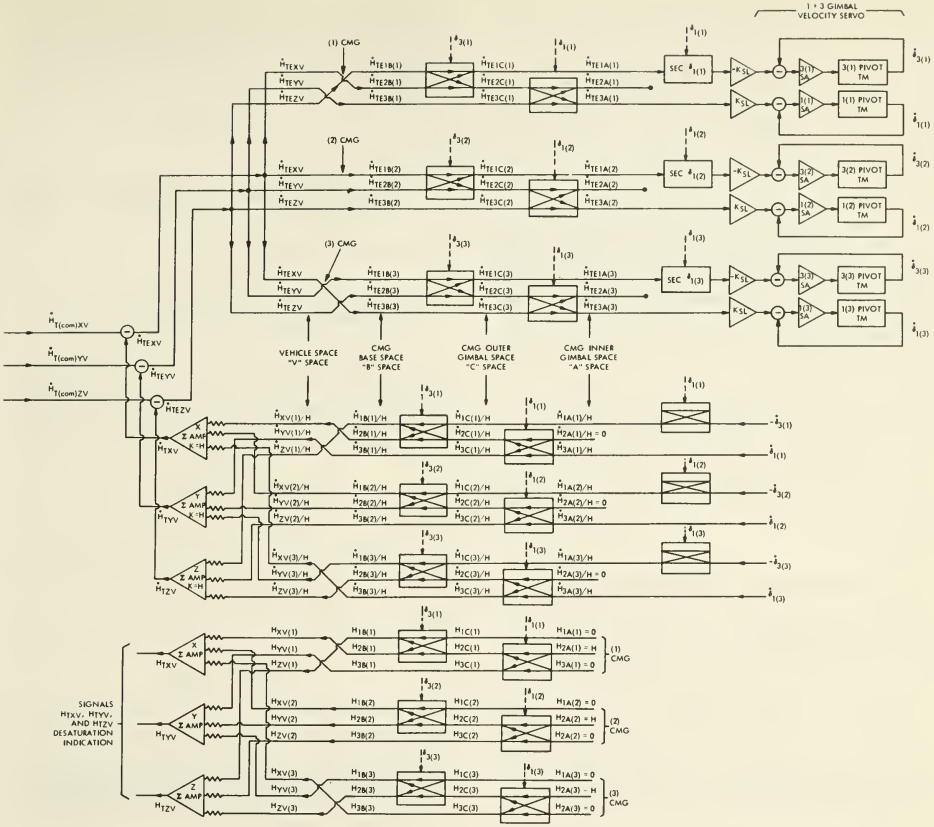


FIGURE 4.4 A BLOCK DIAGRAM OF A CLOSED LOOP $\dot{\bar{H}}_T$ VECTOR CONTROLLER

5.1 THE EQUATIONS OF MOTION OF THE INNER GIMBAL

Newton's law states that, the moment applied to the inner gimbal is equal to the rate of change of the angular momentum vector with respect to the inertial space as shown in equation (5.1).

$$\bar{M} = \frac{d\bar{L}}{dt}_{I\ space} \qquad (5.1)$$

Assume that all cross products of inertia are negligible and also that the only moments generated by product of rotation terms $(W_{IA1A})^2$, $(W_{IC3C})^2$ and $(W_{IA1A} \cdot \tilde{W}_{IC3C})$ are sufficiently large to be considered in the equations of motion of the inner gimbal.

If the above assumptions are made then the equations of motion of the inner gimbal, obtained from expanding equation (5.1) along the $\bar{I}_{1C}$, $\bar{I}_{2C}$ and $\bar{I}_{3C}$ vectors, can be expressed as equations (5.2), (5.3), and (5.4),

$$M_{(C-A)1C} = (J_{A11} + J_D)\,\dot{W}_{IA1A} + H(\sin\delta_1\, W_{IC2C} - \cos\delta_1\, W_{IC3C}) + M_{\epsilon 1} \qquad (5.2)$$

$$M_{(C-A)2C} = -H\sin\delta_1\, W_{IA1A} - \frac{1}{2}\sin 2\delta_1 \cdot \left[(J_{A33} + J_D) - (J_{A22} + J_R)\right] \dot{W}_{IC3C} + M_{\epsilon 2} \qquad (5.3)$$

$$M_{(C-A)3C} = \left[(J_{A33} + J_D)\cos^2\delta_1 + (J_{A22} + J_R)\sin^2\delta_1\right]\dot{W}_{IC3C} - \frac{1}{2}\sin 2\delta_1\left[(J_{A33} + J_D) - (J_{A22} + J_R)\right]\dot{W}_{IC2C} + H\cos\delta_1 W_{IA1A} + M_{\epsilon 3} \quad (5.4)$$

where the vector moment $\bar{M}_{(C-A)}$ is defined as the moment exerted on "A" space by "C" space. The terms $M_{\epsilon 1}$, $M_{\epsilon 2}$, $M_{\epsilon 3}$ in equations (5.2) through (5.4) are products of rotation moments as defined by equation (5.5).

$$M_{\epsilon 1} = \frac{1}{2}\sin 2\delta_1\left[(J_{A33} + J_D) - (J_{A22} + J_R)\right] W^2_{IC3C}$$

$$M_{\epsilon 2} = \left\{(J_{A33} + J_D) - \cos 2\delta_1\left[(J_{A33} + J_D) - (J_{A22} + J_R)\right]\right\} W_{IA1A} W_{IC3C} \quad (5.5)$$

$$M_{\epsilon 3} = -\sin 2\delta_1\left[(J_{A33} + J_D) - (J_{A22} + J_R)\right] W_{IA1A} W_{IC3C}$$

In general, the moments contained in equation (5.5) are small enough at the maximum gimbal rates to also be neglected.

5.2 THE GEARED TORQUE MOTOR EQUATIONS

The inner or (1) and outer or (3) pivots are each constructed so that the inner most frame is driven through a gear train by a D.C. motor which is mounted on the outer frame. The torque motor is shown schematically in Figure 5.1.

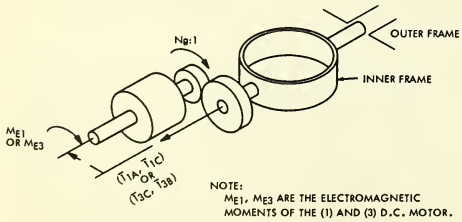


FIGURE 5.1 SCHEMATIC OF (1) OR (3) PIVOT TORQUE MOTOR

Using Newton's law in equation (5.1) the equations of motion of the (1) and (3) pivots at the motor shaft can be expressed as equations (5.6) and (5.7) respectively.

$$N_g M_{E1} + J_{MR}(1 + N_g) N_g \dot{W}_{IC1C} = J_{MR} N_g^2 \dot{W}_{IA1A} + M_{(C-A)1C} \quad (5.6)$$

$$N_g M_{E3} + J_{MR}(1 + N_g) N_g \dot{W}_{IB3B} = (J_{MR} N_g^2 + J_{C33})\dot{W}_{IC3C} + M_{(C-A)3C} \quad (5.7)$$

The moment exerted on the outer gimbal by the (1) pivot gear train or the moment exerted by the outer gimbal on the base along the $\bar{I}_{1C}$ axis is

$$M_{R(C-B)1C} = -M_{(C-A)1C} + J_{MR} N_g \dot{W}_{IA1A} \quad (5.8)$$

The moment exerted on the base by the (3) pivot gear train is

$$M_{R(C-B)3C} = -M_{(C-A)3C} - \left[J_{C33} - N_g J_{MR}\right]\dot{W}_{IC3C} \quad (5.9)$$

where the vector $\bar{M}_{R(C-B)}$ is the reaction moment exerted by the CMG on the spacecraft.

5.3 THE INNER AND OUTER GIMBAL SERVO LOOP KINEMATIC EQUATIONS OF MOTION

The inner or (1) gimbal servo loop kinematic equation of motion can be obtained from equations (5.2) and (5.6) and is expressed as equation (5.10).

$$N_g M_{E1} + J_{MR}(1 + N_g) N_g \dot{W}_{IC1C} = J_{11}\dot{W}_{IA1A} + H\sin\delta_1 W_{IC2C} - H\cos\delta_1 W_{IC3C} + M_{\epsilon 1} \quad (5.10)$$

where the moment of inertia J_{11} is defined as

$$J_{11} = N_g^2 J_{MR} + J_{A11} + J_D . \quad (5.11)$$

The outer or (3) gimbal servo loop kinematic equation of motion can be obtained from equation (5.4) and (5.7) and is expressed as equation (5.12).

$$N_g M_{E3} + J_{MR}(1 + N_g) N_g \dot{W}_{IB3B} = H\cos\delta_1 W_{IA1A} + J_3 \dot{W}_{IC2C} + J_{33}\dot{W}_{IC3C} + M_{\epsilon 3} \quad (5.12)$$

where the moment of inertia terms J_3 and J_{33} are defined in equation (5.13).

$$J_3 = -\frac{1}{2} \sin 2\delta_1 \left\{(J_{A33} + J_D) - (J_{A22} + J_R)\right\} \quad (5.13)$$

$$J_{33} = N_g^2 J_{MR} + J_{C33} + (J_{A33} + J_D) \cos^2 \delta_1 + (J_{A22} + J_R) \sin^2 \delta_1$$

5.4 THE CMG REACTION MOMENT EXERTED ON THE BASE WRITTEN IN "C" SPACE

The reaction moment on the base $M_{R(C-B)1C}$ along the $\bar{l}_{1C}$ vector is obtained from equation (5.2) and (5.8) and is expressed as equation (5.14).

$$M_{R(C-B)1C} = -J_1 \dot{W}_{IA1A} - H \sin\delta_1 W_{IC2C} + H \cos\delta_1 W_{IC3C} - M_{\epsilon 1} \quad (5.14)$$

The reaction moment on the base $M_{R(C-B)2C}$ along the $\bar{l}_{2C}$ vector is the negative value of equation (5.3) and is expressed as equation (5.15).

$$M_{R(C-B)2C} = H \sin\delta_1 W_{IA1A} - J_3 \dot{W}_{IC3C} - M_{\epsilon 2} \quad (5.15)$$

The reaction moment on the base $M_{R(C-B)3C}$ along the $\bar{l}_{3C}$ vector is obtained from equations (5.4) and (5.9) and is expressed as equation (5.16),

$$M_{R(C-B)3C} = -H \cos\delta_1 W_{IA1A} - J_2 \dot{W}_{IC3C} - M_{\epsilon 3} \quad (5.16)$$

where the term $J_3 \dot{W}_{IC2C}$ was omitted from equation (5.16), since J_3 is small compared to the spacecraft moments of inertia.

The moments of inertia J_1 and J_2 are defined as shown in equation (5.17).

$$J_1 = J_{11} - N_g (1 + N_g) J_{MR}$$
$$J_2 = J_{33} - N_g (1 + N_g) J_{MR} \quad (5.17)$$

5.5 THE EQUATIONS OF AN IDEAL CMG

An ideal CMG was defined in section 3.1 as one that has instantaneous servo loops and no inertia reaction moments. Thus from this definition the reaction moments on the base obtained from equations (5.14), (5.15) and (5.16) are modified as

$$M_{R(C-B)1C} = -H \sin\delta_1 W_{IC2C} + H \cos\delta_1 W_{IC3C} \quad (5.18)$$

$$M_{R(C-B)2C} = H \sin\delta_1 W_{IA1A} \quad (5.19)$$

$$M_{R(C-B)3C} = -H \cos\delta_1 W_{IA1A} \quad (5.20)$$

If one considers the exact expression for the reaction moment of an ideal CMG and includes the vehicle body rate then equation (3.7) would be rewritten as equation (5.21).

$$M_R = \bar{W}_{IA} \times \bar{H} \quad (5.21)$$

Expanding equation (5.21) and writing the result in "C" space will provide the same three components $M_{R(C-B)1C}$, $M_{R(C-B)2C}$ and $M_{R(C-B)3C}$ as given in equations (5.18), (5.19) and (5.20). Thus it can be concluded that the math model that has been derived in this section agrees with the formulation used in sections 3 and 4 for the definition of an ideal CMG.

5.6 THE KINEMATIC SIGNAL FLOW DIAGRAM

The kinematic signal flow diagram of the CMG without any electronic control branches is shown in Figure 5.2 and this flow diagram is a representation of equations (5.10), (5.12), (5.14), (5.15) and (5.16), assuming $M_{\epsilon 1}$, $M_{\epsilon 2}$ and $M_{\epsilon 3}$ are negligible. The moments of inertia in Figure 5.2 are defined by equations (5.11), (5.13) and (5.17). Equations (5.14), (5.15), and (5.16) express the reaction moment in "C" space and this moment is transformed from "C" space to "B" space as can be seen by the transformation matrix at the far right of the drawing.

6. THE CMG DYNAMIC CHARACTERISTICS

The electronic and kinematic signal flow graph of a CMG is shown in Figure 6.1, where the solid branches are the kinematic terms as shown in Figure 5.2 and the dashed branches define the electronic configuration of the gimbal velocity servos.

The feedback signals of the gimbal velocity servos are the relative gimbal rates $\dot{\delta}_1$ and $\dot{\delta}_3$ as defined by equation (6.1).

$$\dot{\delta}_1 = W_{IA1A} - W_{IC1C}$$
$$\dot{\delta}_3 = W_{IC3C} - W_{IB3B} \quad (6.1)$$

The node M_{E1} is the sum of two signals, one the amplified error signal $\left\{G_{(1)}(p)\left[\dot{\delta}_{1\,(com)} - \dot{\delta}_1\right]\right\}$ and the other a cross compensation signal

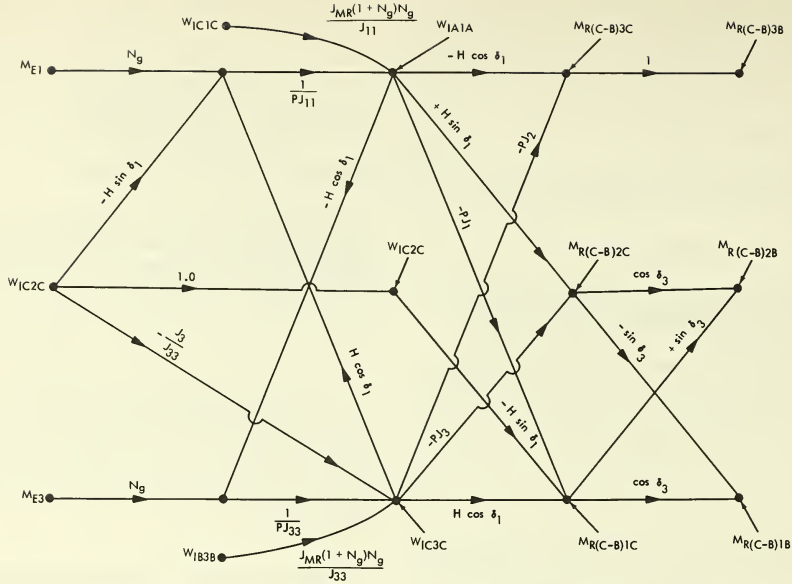


FIGURE 5.2 KINEMATIC SIGNAL FLOW DIAGRAM OF THE CMG

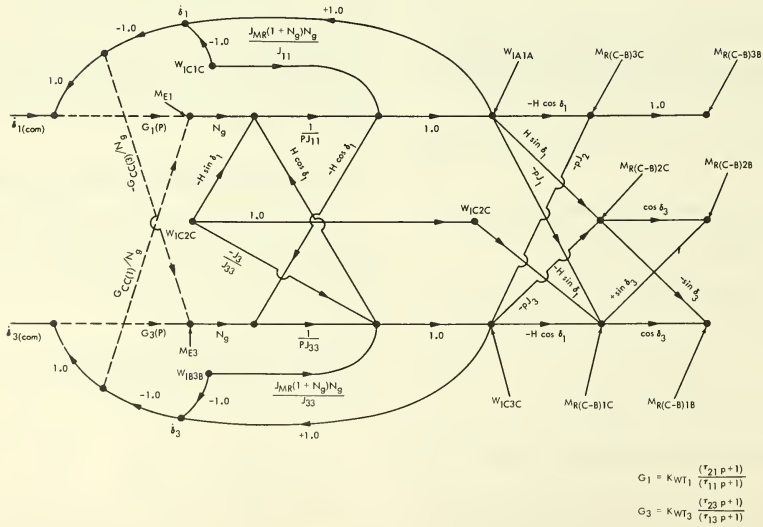


FIGURE 6.1 AN ELECTRONIC AND KINEMATIC SIGNAL FLOW GRAPH OF A CMG

$\left\{\frac{G_{CC(1)}}{N_g}\ \dot{\delta}_3\right\}$. The node M_{E3} is driven in a similar manner to M_{E1} as shown in Figure 6.1. The nodes which define the inertial rates W_{IA1A}, W_{IC2C} and W_{IC3C} are the input variables which generate the CMG reaction moment on the vehicle. The nodes $M_{R(C-B)1C}$, $M_{R(C-B)2C}$, and $M_{R(C-B)3C}$ define the CMG reaction moment on the base in "C" space and the nodes $M_{R(C-B)1B}$, $M_{R(C-B)2B}$, and $M_{R(C-B)3B}$ define the same reaction moment in "B" space.

The input variables to the CMG signal flow diagram are the command relative gimbal rates $\dot{\delta}_{1(com)}$, $\dot{\delta}_{3(com)}$ and the body axis inertial rates W_{IC1C}, W_{IC2C}, and W_{IB3B}.

The left side of the signal flow diagram defines the gimbal velocity servos and the right side defines the method of generating the CMG reaction moment on the base.

6.1 THE CHARACTERISTIC EQUATION OF THE CMG

The two simultaneous equations of motion of the CMG can be obtained from Figure 6.1, where the dependent variables of these equations are $\dot{\delta}_1$ and $\dot{\delta}_3$. These equations are:

$$N_g G_1 \dot{\delta}_{1(com)} - (pJ_1 + N_g G_1) W_{IC1C} - H \sin\delta_1 W_{IC2C} + H \cos\delta_1 W_{IB3B} = (pJ_{11} + N_g G_1)\dot{\delta}_1 - (H\cos\delta_1 - G_{CC(1)})\dot{\delta}_3 \quad (6.2)$$

$$N_g G_3 \dot{\beta}_{3(com)} - H\cos\delta_1 W_{IC1C} - pJ_3 W_{IC2C} - (pJ_2 + N_g G_3) W_{IB3B} = (H\cos\delta_1 - G_{CC(3)})\dot{\delta}_1 + (pJ_{33} + N_g G_3)\dot{\delta}_3 \quad (6.3)$$

where the moments of inertia, J_{11}, J_{33}, J_1, J_2, and J_3, are defined in equations (5.11), (5.13) and (5.17).

The characteristic equation of the CMG can be obtained from equations (6.2) and (6.3) and is given in equation (6.4).

$$\Delta = \left\{(pJ_{11} + N_g G_1)(pJ_{33} + N_g G_3) + (H\cos\delta_1 - G_{CC(1)})(H\cos\delta_1 - G_{CC(3)})\right\} \quad (6.4)$$

In most designs, the torque motor moments of inertia ($N_g^2 J_{MR}$) are considerably larger than the gimbal moments of inertia thus

$$J = J_{11} = J_{33}$$
$$G = G_1 = G_2 \quad (6.5)$$
$$G_{CC} = G_{CC(1)} = G_{CC(3)}$$

and the characteristic equation can be simplified as

$$\Delta = \left\{(pJ + N_g G)^2 + (H\cos\delta_1 - G_{CC})^2\right\}. \quad (6.6)$$

6.2 THE CMG TRANSFER FUNCTIONS

The transfer functions of the CMG for the command input variables $\dot{\delta}_{1(com)}$ and $\dot{\delta}_{3(com)}$ considered one at a time, can be obtained from equations (6.2) and (6.3) and are given in equation (6.7).

$$\frac{\dot{\delta}_1}{\dot{\delta}_{1(com)}} = \frac{W_{IA1A}}{\dot{\delta}_{1(com)}} = \frac{N_g G_1 (pJ_{33} + N_g G_3)}{\Delta}$$

$$\frac{\dot{\delta}_3}{\dot{\delta}_{1(com)}} = \frac{W_{IC3C}}{\dot{\delta}_{1(com)}} = \frac{-N_g G_1 (H\cos\delta_1 - G_{CC(3)})}{\Delta}$$

$$\frac{\dot{\delta}_1}{\dot{\delta}_{3(com)}} = \frac{W_{IA1A}}{\dot{\delta}_{3(com)}} = \frac{N_g G_3 (H\cos\delta_1 - G_{CC(1)})}{\Delta}$$

$$\frac{\dot{\delta}_3}{\dot{\delta}_{3(com)}} = \frac{W_{IC3C}}{\dot{\delta}_{3(com)}} = \frac{N_g G_3 (pJ_{11} + N_g G_1)}{\Delta} \quad (6.7)$$

The other transfer functions for base rotation inputs W_{IC1C}, W_{IC2C}, and W_{IB3B} can also be obtained from equations (6.2) and (6.3) if required.

6.3 THE EXACT EQUATIONS OF THE CMG

The exact equations of the CMG, which form a subroutine for an attitude control system simulation, are given in equations (6.8) through (6.18).

The (1) and (3) servo loop kinematic equations are given in equations (5.10) and (5.12) and are rewritten as equations (6.8) and (6.9).

$$J_{MR}(1 + N_g) N_g \dot{W}_{IC1C} - H \sin\delta_1 W_{IC2C} = J_{11}\dot{W}_{IA1A} - H\cos\delta_1 W_{IC3C} - N_g M_{E1} \quad (6.8)$$

$$J_{MR}(1 + N_g) N_g \dot{W}_{IB3B} - J_3 \dot{W}_{IC2C} = H\cos\delta_1 W_{IA1A} + J_{33}\dot{W}_{IC3C} - N_g M_{E3} \quad (6.9)$$

The (1) and (3) pivot electromagnetic moment equations are obtained from Figure 6.1 and are given in equations (6.10) and (6.11).

$$\tau_{21} K_{WT(1)} \ddot{\delta}_{1(com)} + K_{WT(1)} \dot{\delta}_{1(com)} = \tau_{21} K_{WT(1)} \ddot{\delta}_1 + K_{WT(1)} \dot{\delta}_1 + \tau_{11} \frac{G_{CC(1)}}{N_g} \ddot{\delta}_3 + \frac{G_{CC(1)}}{N_g} \dot{\delta}_3 + \tau_{11} \dot{M}_{E1} + M_{E1} \quad (6.10)$$

$$\tau_{23} K_{WT(3)} \ddot{\delta}_{3(com)} + K_{WT(3)} \dot{\delta}_{3(com)} = - \tau_{13} \frac{G_{CC(3)}}{N_g} \ddot{\delta}_1 - \frac{G_{CC(3)}}{N_g} \dot{\delta}_1 + \tau_{23} K_{WT(3)} \ddot{\delta}_3 + K_{WT(3)} \dot{\delta}_3 + \tau_{13} \dot{M}_{E3} + M_{E3} \quad (6.11)$$

The relative gimbal rates $\dot{\delta}_1$ and $\dot{\delta}_3$ are expressed in equations (6.12) and (6.13).

$$W_{IA1A} = \dot{\delta}_1 + \cos \delta_3 W_{IB1B} + \sin \delta_3 W_{IB2B} \quad (6.12)$$

$$W_{IC3C} = \dot{\delta}_3 + W_{IB3B} \quad (6.13)$$

The body axis rate terms W_{IC1C} and W_{IC2C} are described in equations (6.14), (6.15).

$$W_{IC1C} = \cos \delta_3 W_{IB1B} + \sin \delta_3 W_{IB2B} \quad (6.14)$$

$$W_{IC2C} = -\sin \delta_3 W_{IB1B} + \cos \delta_3 W_{IB2B} \quad (6.15)$$

The reaction moments on the base of the vehicle $M_{R(C-B)1B}$, $M_{R(C-B)2B}$ and $M_{R(C-B)3B}$ are obtained from equations (5.14), (5.15), and (5.16) and when transformed from "C" space to "B" space they can be written as

$$M_{R(C-B)1B} = \cos \delta_3 \left[- J_1 \dot{W}_{IA1A} - H (\sin \delta_1 W_{IC2C} - \cos \delta_1 W_{IC3C}) \right] + \sin \delta_3 \left[- H \sin \delta_1 W_{IA1A} + J_3 \dot{W}_{IC3C} \right] \quad (6.16)$$

$$M_{R(C-B)2B} = \sin \delta_3 \left[- J_1 \dot{W}_{IA1A} - H (\sin \delta_1 W_{IC2C} - \cos \delta_1 W_{IC3C}) \right] + \cos \delta_3 \left[+ H \sin \delta_1 W_{IA1A} - J_3 \dot{W}_{IC3C} \right] \quad (6.17)$$

$$M_{R(C-B)3B} = - H \cos \delta_1 W_{IA1A} - J_2 \dot{W}_{IC3C} \quad (6.18)$$

The moments of inertia, terms J_{11}, J_{33}, J_1, J_2 and J_3, are defined in equations (5.11), (5.13) and (5.17).

6.4 THE EQUATIONS OF AN IDEAL CMG

The attitude control simulation can be simplified by defining an ideal CMG which has instantaneous servo loops and no inertia reaction moments. This simplified model is useful to check the simulation results obtained from the more complex model described in section 6.3.

The instantaneous servo loops require that

$$\dot{\delta}_{1(com)} = \dot{\delta}_1 \qquad \dot{\delta}_{3(com)} = \dot{\delta}_3 \quad (6.19)$$

and the angles δ_1 and δ_3 can be obtained by integrating equation (6.19).

$$\delta_1 = \int_0^t \dot{\delta}_{1(com)} dt + \delta_{1(0+)} \qquad \delta_3 = \int_0^t \dot{\delta}_{3(com)} dt + \delta_{3(0+)} \quad (6.20)$$

The inertial rates W_{IA1A} and W_{IC3C} can be obtained by substituting equation (6.19) in equations (6.12) and (6.13).

$$W_{IA1A} = \dot{\delta}_{1(com)} + \cos \delta_3 W_{IB1B} + \sin \delta_3 W_{IB2B} \quad (6.21)$$

$$W_{IC3C} = \dot{\delta}_{3(com)} + W_{IB3C} \quad (6.22)$$

The body axis rate W_{IC2C} is given by equation (6.15).

$$W_{IC2C} = - \sin \delta_3 W_{IB1B} + \cos \delta_3 W_{IB2B} \quad (6.15)$$

The reaction moments on the base of the vehicle are obtained from equations (5.18), (5.19), (5.20) and when transformed from "C" space to "B" space they can be written as

$$M_{R(C-B)1B} = H\cos\delta_3\left[-\sin\delta_1 W_{IC2C} + \cos\delta_1 W_{IC3C}\right] - H\sin\delta_3\left[\sin\delta_1 W_{IA1A}\right] \quad (6.23)$$

$$M_{R(C-B)2C} = +H\sin\delta_3\left[-\sin\delta_1 W_{IC2C} + \cos\delta_1 W_{IC3C}\right] + H\cos\delta_3\left[\sin\delta_1 W_{IA1A}\right] \quad (6.24)$$

$$M_{R(C-B)3C} = -H\cos\delta_1 W_{IA1A} \quad (6.25)$$

If the initial conditions δ_1 (0+) and δ_3 (0+) are zero, then considerable simplification occurs in the math model defined by equations (6.21) through (6.25).

6.5 THE ATM/CMG RATE SERVO CHARACTERISTICS

A signal flow graph of the (1) and (3) pivot gimbal angular velocity servos is shown in Figure 6.1. The electronic transfer functions are shown as dashed lines and the kinematic transfer functions are solid lines. The numerical parameters for all transfer functions are displayed in Table 6.1. The forward path electronic functions, $G_1(P)$ and $G_3(P)$ are lead-lag compensation networks as shown in Table 6.1. The cross-feed transfer functions, Gcc(1) and Gcc(3), determine the amount of decoupling between the (1) and (3) pivot servos. Ideally, the cross-feed transfer functions are equal to Hcos δ_1, which will completely decouple the two servo loops. However, this approach requires the use of two extra resolvers on the (1) pivot and it can be shown that a constant cross-feed transfer function decouples the loops adequately for an inner (δ_1) gimbal freedom of ±70°.

H	=	2000 ft-lb-sec
J_{11}	=	17.66 ft-lb-sec^2
J_{33}	=	21.05 ft-lb-sec^2
N_g	=	56
$G_1(P)$	=	$1575\ \frac{(5P+1)}{(.05P+1)}$ ft-lb-sec/rad
$G_3(P)$	=	$1925\ \frac{(5P+1)}{(.05P+1)}$ ft-lb-sec/rad
$G_{cc(1)}$	=	$G_{cc(2)}$ = 1414 ft-lb-sec

TABLE 6.1 ATM/CMG PARAMETER VALUES

Figure 6.2 shows the closed loop frequency response of $\dot{\delta}_1/\dot{\delta}_1$ (com), as defined by equations 6.4 and (6.7), with no cross-feed transfer function for inner gimbal angles of 0° and 70°. The numerical values other than the cross-feed terms are given by Table 6.1.

Figure 6.3 shows the closed loop frequency response with a constant cross-feed transfer function of Hcos45° (e.g. Gcc(1)/N_g = Gcc(3)/N_g = 25.25 ft., lb., sec.), for inner gimbal angles of 0, 45, and 70 degrees.

A comparison of Figures 6.2 and 6.3 shows that the constant cross-feed transfer function is desirable since it will reduce the variations in the $\dot{\delta}_1/\dot{\delta}_1$ (com) transfer function so that it can almost be considered to be independent of the (1) pivot angle. Also, the constant cross-feed transfer function provides a nearly constant bandwidth of 40 radians/second whereas the absence of the cross-feed transfer function exhibits a bandwidth between 8 and 40 radians/second depending on the (1) pivot angle.

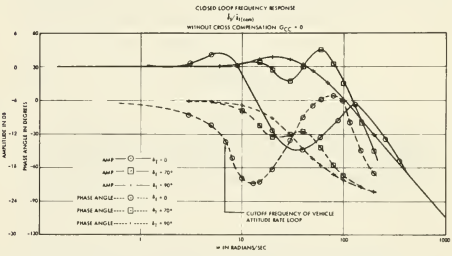


FIGURE 6.2 CLOSED LOOP FREQUENCY RESPONSE WITHOUT CROSS COMPENSATION

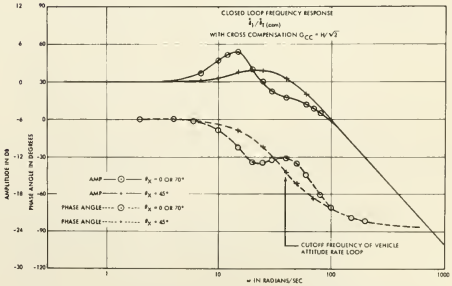


FIGURE 6.3 CLOSED LOOP FREQUENCY RESPONSE WITH CROSS COMPENSATION

67-589

7. CONCLUSIONS

The use of momentum exchange is desirable for attitude control of space vehicles, particularly where accurate, continuous control of large spacecraft over long durations is required. The efficiencies and dynamic characteristics of clusters of two or more CMGs make the use of momentum exchange a practical method of control of spacecraft.

The particular cluster to be used will depend on tradeoff considerations which involve size, weight, power, allowable control law complexity, momentum utilization, degree of redundancy, etc. For the ATM application the SIX-PAC[1] configuration exhibits the most desirable characteristics because of its excellent (100%) redundancy, momentum utilization and yet moderate size, weight, and power characteristics.

Extensive analyses of the control law for the ATM/CMG control system indicates that the closed loop position $\bar{H}$-vector, or closed loop rate $\dot{\bar{H}}$ vector, control laws are the best choice for the CMG control laws. These control laws exhibit reasonably constant gain and minimum cross-coupling. From an implementation standpoint, the closed loop $\bar{H}$ vector control law requires less complexity and provides for a maximum utilization of the hardware.

REFERENCES

[1] Kurzhals, Peter, R. and Grantham, Carolyn: A System for Inertial Experiment Pointing and Attitude Control, proposed NASA TR, dated May, 1965.

[2] Bendix Corporation and Douglas Aircraft Corporation: MORL Mechanical Systems, Volume XIV, Stabilization and Control, Douglas Report No. SM-46806, prepared under NASA Contract NAS 1-3612, dated September 1964.

[3] Seltzer, S., Schultz, D., and Chubb, W.: Attitude Control and Precision Pointing of the Apollo Telescope Mount, AIAA Conference Paper, dated August, 1967.

ACKNOWLEDGMENT

The writers would like to acknowledge the technical assistance of the following people in the background material and preparation of this paper:

Dr. Peter R. Kurzhals

Spacecraft Systems Branch
Applied Material and Physics Division
Langley Research Center, Hampton, Va.

Dr. Sherman Seltzer, Branch Chief

Guidance and Control Systems Branch
Guidance and Control Division
Astrionics Laboratory
Marshall Space Flight Center, Huntsville, Alabama

David Schultz

Guidance and Control Systems Branch
Guidance and Control Division
Astrionics Laboratory
Marshall Space Flight Center, Huntsville, Alabama

Carl Mandel, Chief

Inertial Sensors and Stabilizers Division
Astrionics Laboratory
Marshall Space Flight Center, Huntsville, Alabama

The authors are indebted to Mr. Herman Thomason (Deputy Chief of the Inertial Sensor and Stabilizer Division, The Astrionics Laboratory) for his collaboration in the preparation of the material that appears in sections 3., 5. and 6.

APPENDIX

A DEFINITION OF SPACES, TRANSFORMATION MATRICES AND SYMBOLS

All spaces, transformation matrices and symbols were defined as used throughout the text. The definitions contained in this section are the same as those defined previously and are compiled for easy reference.

A DEFINITION OF SPACES

There are eleven different spaces used through the discussion, nine of these spaces are assigned to the three CMG and the other two spaces are vehicle space and inertial space. These spaces are:

"$A_{(j)}$" space or the inner gimbal of the "j^{th}" CMG which is described by the coordinate system $\bar{I}_{1A(j)}$, $\bar{I}_{2A(j)}$ and $\bar{I}_{3A(j)}$ for j = 1, 2, 3.

"$C_{(j)}$" space or the outer gimbal of the "j^{th}" CMG which is described by the coordinate system $\bar{I}_{1C(j)}$, $\bar{I}_{2C(j)}$ and $\bar{I}_{3C(j)}$ for j = 1, 2, 3.

"$B_{(j)}$" space or the base of the "j^{th}" CMG which is described by the coordinate system $\bar{I}_{1B(j)}$, $\bar{I}_{2B(j)}$ and $\bar{I}_{3B(j)}$ for j = 1, 2, 3.

"V" space or vehicle space which is described by the coordinate system $\bar{I}_{XV}$, $\bar{I}_{YV}$ and $\bar{I}_{ZV}$.

"I" space or inertial space which is described by the coordinate system $\bar{I}_{XI}$, $\bar{I}_{YI}$ and $\bar{I}_{ZI}$.

A DEFINITION OF TRANSFORMATION MATRICES

There are six different transformation matrices used throughout the discussion and their definitions are:

"$A_{(j)}$" to "$B_{(j)}$" space	-	equation (1.3)
"$A_{(j)}$" to "$C_{(j)}$" space	-	equation (1.4)
"$C_{(j)}$" to "$B_{(j)}$" space	-	equation (1.5)
"$B_{(1)}$" to "V" space	-	equation (2.5)
"$B_{(2)}$" to "V" space	-	equation (2.6)
"$B_{(3)}$" to "V" space	-	equation (2.7)

A DEFINITION OF SYMBOLS

The CMG gimbal angles $\delta_{1(j)}$ and $\delta_{3(j)}$ are defined as

$\delta_{1(j)}$ ≙ The (1) pivot angle of the (j) CMG - see Figure 1.4.

$\delta_{3(j)}$ ≙ The (3) pivot angle of the (j) CMG - see Figure 1.5.

$\dot{\delta}_{1(j)}$, $\dot{\delta}_{3(j)}$ ≙ The time rate of change of angles $\delta_{1(j)}$, $\delta_{3(j)}$.

$\dot{\delta}_{1(com)(j)}$, $\dot{\delta}_{3(com)(j)}$ ≙ The (1) and (3) pivot gimbal rate servo inputs or command of the "j^{th}" CMG.

Note: In sections 5. and 6., the (j) subscript has been omitted since the discussion concerns one CMG.

The angular momentum symbols $\bar{H}_{T(com)}$, $\bar{H}_T$, $\bar{H}_{TE}$, $\bar{H}_{(j)}$ are defined as:

$\bar{H}_{T(com)}$ ≙ The time integral of the attitude control system moment command (see figure 3.1)

$\bar{H}_T$ ≙ The total momentum vector of the (1), (2) and (3) CMGs (see equation (3.4)).

$\bar{H}_{TE}$ ≙ The error signal between $\bar{H}_{T(com)}$ and $\bar{H}_T$ (see equation (3.9)).

$\bar{H}_{(j)}$ ≙ The angular momentum vector of the "j^{th}" CMG.

$H_{T(com)XV}$, $H_{T(com)YV}$, $H_{T(com)ZV}$,
H_{TXV}, H_{TYV}, H_{TZV},
H_{TEXV}, H_{TEYV}, H_{TEZV}

≙ The X, Y, Z components of vectors $\bar{H}_{T(com)}$, $\bar{H}_T$ and $\bar{H}_{TE}$ in vehicle space.

$H_{1A(j)}$, $H_{2A(j)}$, $H_{3A(j)}$,
$H_{TE1A(j)}$, $H_{TE2A(j)}$, $H_{TE3A(j)}$

≙ The components of vectors $\bar{H}_{(j)}$ and $\bar{H}_{TE}$ along coordinates $\bar{I}_{1A(j)}$, $\bar{I}_{2A(j)}$ and $\bar{I}_{3A(j)}$.

$H_{1C(j)}$, $H_{2C(j)}$, $H_{3C(j)}$,
$H_{TE1C(j)}$, $H_{TE2C(j)}$, $H_{TE3C(j)}$

≙ The components of vector $\bar{H}_{(j)}$ and $\bar{H}_{TE}$ along coordinates $\bar{I}_{1C(j)}$, $\bar{I}_{2C(j)}$ and $\bar{I}_{3C(j)}$.

$H_{1B(j)}$, $H_{2B(j)}$, $H_{3B(j)}$,
$H_{TE1B(j)}$, $H_{TE2B(j)}$, $H_{TE3B(j)}$

≙ The components of vector $\bar{H}_{(j)}$ and $\bar{H}_{TE}$ along coordinates $\bar{I}_{1B(j)}$, $\bar{I}_{2B(j)}$ and $\bar{I}_{3B(j)}$.

The time rate of change of angular momentum symbols $\dot{\bar{H}}_{T(com)}$ and $\dot{\bar{H}}_T$ are defined as:

$\dot{\bar{H}}_{T(com)}$ ≙ The time rate of change of vector $\bar{H}_{T(com)}$ with respect to "V" space, $\dot{\bar{H}}_{T(com)}$ is proportional to the linear combination of attitude error and vehicle body axis rate.

$\dot{\bar{H}}_T$ ≜ The time rate of change of vector $\bar{H}_T$ with respect to "V" space.

$\dot{\bar{H}}_{TE}$ ≜ The error signal between $\dot{\bar{H}}_{T(com)}$ and $\dot{\bar{H}}_T$ (see equation (4.7)).

The angular velocity vectors $\bar{W}_{VA(com)}$, $\bar{W}_{IA}$, $\bar{W}_{IC}$, and $\bar{W}_{IB}$ are defined as:

$\bar{W}_{VA(com)(j)}$ ≜ The desired inner gimbal rate of the "j^{th}" CMG with respect to vehicle space.

$\bar{W}_{VA(j)}$ ≜ The actual inner gimbal rate of the "j^{th}" CMG with respect to vehicle space.

$\bar{W}_{IA}$ ≜ The CMG inner gimbal rate with respect to inertial space.

$\bar{W}_{IC}$ ≜ The CMG outer gimbal rate with respect to inertial space.

$\bar{W}_{IB}$ ≜ The CMG base rate with respect to inertial space.

Note: The subscript (j) has been omitted from the terms $\bar{W}_{IA}$, $\bar{W}_{IC}$, $\bar{W}_{IB}$, since these symbols are used in sections 5. and 6., and the discussion concerns only one CMG.

W_{IA1A} ≜ The $\bar{1}_{1A}$ component of the vector $\bar{W}_{IA}$.

W_{IC1C}, W_{IC2C}, W_{IC3C} ≜ The components of vectors $\bar{W}_{IC}$ along coordinates $\bar{1}_{1C}$, $\bar{1}_{2C}$ and $\bar{1}_{3C}$.

W_{IB3B} ≜ The $\bar{1}_{3B}$ component of the vector $\bar{W}_{IB}$.

The moments $\bar{M}_R$, $M_{R(C-B)}$, $\bar{M}_{R(C-A)}$, M_{E1}, M_{E3}, $M_{\epsilon 1}$, $M_{\epsilon 2}$ and $M_{\epsilon 3}$ are defined as follows:

$\bar{M}_R = \bar{M}_{R(C-B)}$ ≜ The reaction moment of the CMG on the base of the vehicle.

$\bar{M}_{(C-A)}$ ≜ The moment exerted on the inner gimbal by the outer gimbal.

M_{E1}, M_{E3} ≜ The electromagnetic moments of the (1) and (3) pivot D.C. motors.

$M_{\epsilon 1}$, $M_{\epsilon 2}$, $M_{\epsilon 3}$ ≜ The kinematic moments obtained from product of rotation terms (see equation (5.5)).

$M_{R(C-B)1C}$, $M_{R(C-B)2C}$, $M_{R(C-B)3C}$

≜ The components of the vector $\bar{M}_{R(C-B)}$ along coordinates $\bar{1}_{1C}$, $\bar{1}_{2C}$ and $\bar{1}_{3C}$.

$M_{R(C-B)1B}$, $M_{R(C-B)2B}$, $M_{R(C-B)3B}$

≜ The components of the vector $\bar{M}_{R(C-B)}$ along coordinates $\bar{1}_{1B}$, $\bar{1}_{2B}$ and $\bar{1}_{3B}$.

The moments of inertia J_R, J_D, J_{A11}, J_{A22}, J_{A33}, J_{C33}, J_{MR}, J_{11}, J_{33}, J_1, J_2, J_3, are defined as

J_R, J_D, J_{A11}, J_{A22}, J_{A33}, J_{C33}, J_{MR} ≜ Are polar moments of inertia as defined in section 5.0.

J_{11} ≜ A polar moment of inertia (see equation (5.11)).

J_3, J_{33} ≜ Polar moments of inertia (see equation (5.13)).

J_1, J_2 ≜ Polar moments of inertia (see equation (5.17)).

H ≜ The angular momentum of the (1), (2) and (3) CMG.

N_g ≜ The (1) and (3) pivot gear ratios.

The (1) and (3) pivot gimbal servo amplifier functions are

$$G_1(p) = K_{WT(1)} \frac{\{\tau_{21} p + 1\}}{\{\tau_{11} p + 1\}} \quad \text{(1) pivot}$$

$$G_3(p) = K_{WT(3)} \frac{\{\tau_{23} p + 1\}}{\{\tau_{13} p + 1\}} \quad \text{(3) pivot}$$

$K_{WT(1)}$, $K_{WT(3)}$ ≜ The D.C. gain of the (1) and (3) pivot servo amps.

τ_{11}, τ_{13} ≜ The larger time constant of the (1) and (3) pivot servo amplifier lag network.

τ_{21}, τ_{23} ≜ The smaller time constant of the (1) and (3) pivot servo amplifier lag network.

$G_{CC(1)}$, $G_{CC(3)}$ ≜ The electronic cross compensation from (3) pivot to (1) pivot and (1) pivot to (3) pivot respectively.

No. 67-590

RESULTS OF EXPERIMENTAL STUDIES IN THE ACTIVE DAMPING OF FREE-ROTOR GYROSCOPES

by

BRADFORD W. PARKINSON
Aerospace Research Pilots School
Edwards Air Force Base, California
and
BENJAMIN O. LANGE
Stanford University
Stanford, California

AIAA Paper
No. 67-590

AIAA Guidance, Control and Flight Dynamics Conference
HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

-- NOTES --

RESULTS OF EXPERIMENTAL STUDIES IN THE ACTIVE DAMPING OF FREE-ROTOR GYROSCOPES*

Bradford W. Parkinson†
Major, U. S. Air Force
Aerospace Research Pilots School
Edwards Air Force Base, California

Benjamin O. Lange ††
Assistant Professor
Department of Aeronautics and Astronautics
Stanford University
Stanford, California

Abstract

In a free-rotor gyro the nearly spherical gyro-rotor is not mechanically constrained to rotate about a particular axis. Therefore, it may be initially spun up about any rotor-fixed axis.

As seen in the rotor, the subsequent motion of the angular velocity vector after spin-up is the well-known classical polhode motion about either the maximum or the minimum principal axes of inertia. The maximum axis of inertia is the desired spin-axis direction because it is the only stable free-rotor spin direction in the presence of minute energy losses during flexures of the rotor. Consequently, the rotor-fixed markings used for spin-axis readout of many of these gyros are referenced to the maximum axis of inertia. The principle delays in activating typical gyros are (1) waiting for a passive DC field to damp the gyro spin-axis into the maximum axis of inertia of the rotor after it has been spun up (for current designs this requires typically five times the spin-up time) and (2) waiting for the thermal transient (created by the damping field) to subside.

This paper reports the experimental studies of a method of active damping which was described and analyzed in an earlier paper. Both coarse and fine active damping were mechanized in an air-bearing gyro, and the experimental damping characteristics are compared with the results of a theoretical analysis. The experiments agreed very well with the theory and show that damping to an accuracy of a few arc seconds can be completed in the same amount of time required to bring the rotor to operating speed. Furthermore, the rotor can be damped in so that the same side is always up. This eliminates the necessity for the choice between two error models for the gyro or multiple start-up attempts until the correct side is up.

INTRODUCTION

As explained in Ref. 1, free-rotor gyros require a spin-axis direction along the axis of maximum moment of inertia in the rotor. This constraint arises in gyros which depend upon rotor-fixed markings for readout. Given enough time, the mechanism of internal energy dissipation will cause the gyro spin direction to damp into this rotor-fixed axis, but the time constants involved may be from days to months, depending on rotor material, moment-of-inertia ratios, size, spin rate, and other factors.

For most applications, these times are unacceptable; some scheme of artificially damping the spin-axis to the maximum axis of inertia must be used. The passive method of applying a strong DC field parallel to the spin-axis suffers because of the temperature transients it creates, and because the time required in current designs is typically about five times the time required just to bring the rotor up to speed. An active damping scheme, which operates simultaneously with rotor spin-up, has been developed that can reduce these damping times to less than the time required for spin-up.

Active damping is defined as making the torque axis of an erection field be a function of some component of rotor angular velocity in order to re-orient the spin direction in the rotor while leaving its direction in a laboratory or inertial frame basically unaltered.

The problem of active damping can be divided into two parts: coarse damping for initial spin velocity vectors which are as much as 180 degrees from the desired spin-axis, and fine damping which achieves residual polhodes of only a few arc seconds. Ref. 1 describes the mathematical analysis of a solution to the coarse damping problem which we call "hemispheric torquing". The essential part of this scheme is to apply an additional torque perpendicular to the spin-up torque whose average in inertial space is zero; but, because it switches in synchronization with the rotor rotation, it has a net effect when viewed in the rotor frame of reference. This net torque can be considered as a rectification in the rotor's coordinates. In effect, the angular rate of the rotor remains fixed in the laboratory while the rotor precesses beneath it to the desired axis.

A mathematical analysis of a solution to the fine-damping problem is contained in Ref. 2. This solution makes use of the fact that a spin-axis readout such as the Honeywell D-Pattern system contains (in addition to the usual spin axis-direction information) a term in its output whose amplitude is proportional to the polhode amplitude and which varies sinusoidally at polhode frequency. This term may be separated from the spin-axis-direction information by a low pass filter which cuts off just above the polhode frequency and is used to command extra body fixed moments which damp out the polhodes.

*The research reported here is part of a program supported at Stanford University by the U.S. Air Force under Contract AF33(615)-1411 from the Air Force Avionics Laboratory.

†Major, U.S.A.F. This work is part of Major Parkinson's doctoral thesis. Currently assigned to the staff of the Aerospace Research Pilots School, Edwards Air Force Base, California.

††Assistant Professor, Department of Aeronautics and Astronautics, Stanford University, Stanford, California.

The experiments show that beginning with a worst-case initial condition where the rotor is turned 180° away from the desired spin-axis, active damping can be completed in the same time that is required to bring the rotor up to speed and achieve residual polhodes of about 5 arc seconds amplitude.

DESCRIPTION OF THE APPARATUS

1. Gyro and Rotor

The gyro is shown in Fig.1. Three orthogonal air-bearing pads are located in the lower three faces of the case cube. Each face of the case contains a torquing coil, making three orthogonal pairs in all. Three orthogonal axes designated X, Y, and Z are defined to lie parallel to the three long sensors shown attached to the gyro in the figure. The z-axis comes out of the figure to the left, the x-axis comes out of the figure to the right, and the y-axis goes into the figure and to the left. The z-axis is the nominal spin-axis of the gyro in the laboratory reference frame. The two coils designated X and Y had a 5-kc sine wave applied in quadrature; they were designated as spin-up coils and were used to generate a spin-up torque along the case Z axis. Each also created one phase of the spinning field that generated torques along the X and Y axes. The other phase was supplied by currents in the Z coils. The X and Y torques were used for hemispheric torquing and spin-axis control. Current in the Z coil was switched and phased to provide these torques as called for by the control law.

Beryllium-copper was selected as the rotor material in order to minimize the possibility of permanent magnetism, which causes body-fixed moments. The rotor diameter was 2.0000 in., ±0.0001 in.; sphericity was better than 25 μin. The measured moment-of-inertia difference ratio was 3×10^{-4}. † The rotor is illustrated in Fig. 2. Notice the dull and bright "hemispheres" and the serrated markings of the dentate "pattern". The axis of maximum inertia is perpendicular to the average equator.

The gyro was mounted on a 3/4-in. aluminum plate secured to a concrete pier with three 3/4-in. stud bolts as shown in Fig. 3. Nitrogen was fed to the support pads from a regulated line which passed through holes in the aluminum plate.

The gyro induction motor is excited with current at 5 KHz. Part of the z-axis current is switched 180° in phase according to whether the z-sensor detects a dull or a bright hemisphere and a frequency of 5 KHz was dictated by the number of teeth in the dentate pattern, shown in Fig. 2.

Eleven teeth call for torque switching at 22 times per rotor revolution. At 3000 rpm this 5-kc signal reverses roughly every three cycles. Any lower excitation frequency would place a very stringent requirement on the power amplifiers to effect phase reversal in the highly inductive coils. The maximum rotor spin speed is under 100 cps; therefore the motor slip [Ref.3] is essentially constant for all operating speeds. The model of Section 221 of Ref. 3 is adequate for the experimental verification; in an air-bearing gyro, any uncertainty in the motor model is dwarfed by the uncertainty in air-bearing torques. The torque produced by the spin-up motor is due to the currents $i_x = i_y$ being in quadrature and is given by

$$M_m = \frac{V^2 R_2'}{2\Omega_s}\left[\frac{S_m(1+k)^2}{[R_1 S_m + R_2']^2 + \chi^2 S_m^2} - \frac{(2-S_m)(1-k)^2}{[R_1(2-S_m)+R_2']^2 + \chi^2(2-S_m)^2}\right]$$

where V = voltage applied to the x or y coil (volts)
Ω_s = frequency of excitation (rad/sec)
S_m = slip (percentage of Ω_s)
R_2' = rotor resistance referred to the stator (ohms)
R_1 = coil resistance (ohms)
χ = reactance of the coil including the effect of the rotor (ohms)
k = 1, V_{zy}/V, and V_{zx}/V for Z, Y, and X torques respectively
V_{zx}, V_{zy} = Z coil voltage in quadrature with the Y and X coil respectively.

For the spin rates considered, $S_m \cong 1$ and, since the coil reactances are nearly equal, $V_x/i_x = V_y/i_y = V_{zy}/i_{zy} = V_{zx}/i_{zx}$. Therefore the equations for torque become

$$M_z = K_m i_x^2, \quad M_y = K_m i_x i_{zy}, \quad M_x = K_m i_y i_{zx} \tag{1}$$

where M_z, M_y, M_x = case Z, Y, and X axis torques respectively
K_m = torque/(amp)^2 (determined experimentally)
i_x = balanced X or Y coil current
i_{zy} = Z coil current in quadrature with X coil
i_{zx} = Z coil current in quadrature with Y coil

The constant K_m has been separated from other air-bearing effects by assuming a model:

$$M_z = K_m (i_x)^2 - L(W_r - W_o) \tag{2}$$

where L = coefficient of air drag resistance
W_r = rotor angular rate
W_o = effective zero current angular rate of the rotor due to gyro air torques

The experimentally determined air-bearing motor step-response curves are shown in Fig. 4. The coefficient K_m for the experiment was determined to be

$$K_m = 0.059 \pm 0.005 \ \frac{\text{rad/sec}^2}{(\text{rms amp})^2}$$

The coefficient 1/L was experimentally found as

$$1/L = 255 \pm 10 \text{ sec}$$

2. Dentate Pattern and Active Caging

The question of how to mark the rotor in order to implement active damping is a difficult one due to the number of constraints. Any marking scheme must allow the hemispheres to be uniquely distinguished (hemisphere will be used in a general

† It seems strange that a rotor so precisely machined would have a difference ratio this large. An unconfirmed conjecture is that the rotor was initially stamped, creating anisotropic internal stresses and a lack of homogeneity. A homogeneous rotor of this quality would probably have difference ratios of less than 10^{-5}.

sense: it may not refer to a great-circle dividing line). It must also, in the case of D-pattern use, incorporate the D-pattern and phase marking; it should allow for some means of actively controlling the inertial spin-axis during hemispheric torquing; and finally it should be permanent, yet have negligible effect on the inertial properties of the rotor. These problems can be divided into two parts: geometry and method of application. First let us consider geometry.

Fig. 5 illustrates a pattern which we shall call dentate. It has a number of desirable features from a geometric viewpoint. The equally spaced teeth are formed from great circles crossing the equator at 45°. They divide the rotor roughly into northern and southern hemispheres. As the rotor damps (under the action of hemispheric torquing), it reaches a point at which the locus of the hemisphere sensor in the rotor no longer clears the points of the teeth. This condition leads quite naturally to the saturation type of control law which was postulated in the mathematical description for the digital simulation in Ref. 1. The point is illustrated by Figs. 6 and 7. For the discussion, the rule of hemispheric torquing is applied; that is, if the sensor is opposite the "northern" surface the X inertial torque is positive, and conversely.

If we define E (called the co-elevation angle) to be the angle between the angular velocity vector of the rotor and the desired spin-axis in the rotor, then for large E the average body X-axis torque is 0.636. As E decreases, the hemisphere sensor grazes the tops of the teeth and there is a corresponding decrease in average torque amplitude. Obviously when the sensor is traversing the equator, there is no average effect at all. Note that this average is exactly the equivalent of the term $K_2 \sin(\theta^*)$ in the Kryloff-Bogoliuboff equations used in the analysis of Ref. 1.

When the average torque amplitude goes to zero, it corresponds to zero input into the KB equations. Relative to E, the saturation-function characteristics will vary according to the number of teeth equally spaced around the rotor. If the great-circle locus of the hemisphere sensor is determined using spherical trigonometry, the radian distance h_σ of the locus from the equator is

$$h_\sigma = \sin^{-1}[\sin E \sin(Wt + \alpha_\sigma)]$$

where W is the angular rate of the rotor and α_σ is an arbitrary initial phase. This situation is illustrated in Fig. 8.

Similarly,

$$\phi_\sigma = \tan^{-1}[\tan(Wt + \alpha_\sigma)\cos E]$$

where ϕ_σ is the distance along the rotor equator (radians), shown in Fig. 8. The height of the marker above the equator (Fig. 9) may be shown to be

$$h_\mu = \tan^{-1}\left[\sin\left(\frac{\pi}{2n} - |Q|\right)\right]$$

where $Q \triangleq [(\phi_\sigma - \beta_\sigma) \bmod (2\pi/n)] - (\pi/n)$

n = number of teeth

β_σ = angle from the intersection of sensor locus and equator to the first down tooth

The average torque component along the rotor X-axis is given by

$$\frac{\langle M_{xb}\rangle(\beta_\sigma, E, n)}{M_{max}\ \frac{W}{2\pi}} = \int_0^{2\pi/W} \sin(Wt+\alpha_\sigma)\ \mathrm{sgn}[h_\sigma(E,t) - h_\mu(\beta_\sigma,t)]\,dt$$

The quantity $\langle M_{xb}\rangle$ has been evaluated for $\beta_\sigma = (2\pi/n)(j/8)$ where j = 1, 2,...., 7, and averaged over this set of the β_σ for a fixed value of E. Taking $M_{max} = 1$ the results are plotted in Fig. 10 as average torque $\left(\langle M_{xb}\rangle\right)$ vs. co-elevation (E) for various numbers of teeth (n). The linear range of the saturation function vs. (1/teeth) is plotted in Fig. 11.

The value of having the saturation function rather than the pure on-off control is that it insures a stable null without limit cycles when the erection torque is properly adjusted and active spin-caging control is used.

Active spin-caging control is achieved in the following manner. First we will only consider odd numbers of teeth. An odd number of teeth has the advantage that any great-circle locus of the sensor will have equal times in the "north" and "south". Therefore the inertial X-axis torque will average to zero. However, if the spin-axis shifts slightly to or away from being perpendicular to the plane of the hemisphere sensor, while remaining fixed in the rotor, it causes the inertial average to become non-zero.

The expression for this condition is

$$\frac{\langle M_{xi}\rangle(\beta_\sigma, E, n, \delta)}{M_{max}\ \frac{W}{2\pi}} = \int_0^{2\pi/W} \mathrm{sgn}[h_\sigma(E,t,\delta) - h_\mu(\beta_\sigma,t)]\,dt$$

where the first-order approximation to $h_\sigma(E,t,\delta)$ is

$$h_\sigma = \sin^{-1}[\sin E \sin(Wt + \alpha_\sigma)] + \delta$$

where δ (measured in radians) is the component of misalignment of spin-axis about the axis* of the y-sensor, and α_σ may be taken as zero with no loss of generality in the averaging. For n = 11, the function $\langle M_{xi}\rangle(\beta_\sigma, E, n, \delta)$ is plotted vs. δ (after averaging over the same set of β_σ as before) for various E in Fig. 12.

For values of $E \geq 30°$, the teeth exert little influence on the average output, and they may be replaced with a single great circle to estimate the effect of changes in spin-axis position on the average output. The situation is shown in Fig. 13.

The average change in output (to first-order in δ) may be shown to be

$$\langle \Delta_{out} \rangle = \frac{2\eta}{2\pi} M_{max} = \frac{\delta \cot E}{\pi} M_{max} \qquad (3)$$

where η is the angle defined in Fig. 13.

The method of spin-caging control, when damping in the large, is to use the average hemispheric

* The component about the x-sensor axis does not change the average output of the x-sensor.

torque as a measure of spin-axis misalignment to alter the direction of the spin-up field and thus reduce this uncertainty. If only the X-axis control and Z-axis spin-up torques are considered, the character of this control will be as revealed in Fig. 14. The Y-axis is controlled in an analogous manner. Thus, the addition of spin-caging control shifts the average torque direction but leaves the hemispheric torquing mechanism unchanged.

Unfortunately, Eq. (3) shows that the effect of misalignment changes sign for angles greater than $\pi/2$, so a spin-caging control with constant gain would have a destabilizing effect in this region. Use of the spin-axis sensor, located on the Z case axis, solves this problem. The sign of the spin-caging control gain is changed according to whether this sensor, on the average, sees the northern or southern hemisphere. This poses one additional problem, since the D-pattern will lead to improper decisions when the spin-axis is in it and under the spin-axis sensor. The solution to this problem is to use the fact that this condition implies that X and Y sensors are sensing a continuous dentate pattern. The condition that the dentate pattern is continuously present is used logically to override the decision based on Z-sensor output. Continuous dentate pattern is implied by the existence of the rotor phase times T_j. (See the next paragraph and Fig. 18.) The control is diagramed for the X channel in Fig. 15. A saturation function is included in the caging loop to allow the caging torque to exceed the hemispheric torque but not saturate the amplifiers.

An 11-tooth dentate pattern was selected for the model. It represents a compromise between torque-switching characteristics and saturation-function linear range. The pattern was modified as shown in the "polar map" of Fig. 16a; this allowed the event times to be determined logically independent of rotor spin speed. The logic scheme essentially counted state changes of one sensor during a continuous state on the other. If the number of state changes exceeded two, a phase time had occurred. Taken into consideration in this design modification were (1) the goal of symmetry of the pattern to insure that the spin-caging control is unbiased, and (2) the reduction of hemispheric-torquing biases by insuring that the modification does not bias any quadrant of the equator which would lead to body-rectified torques and a balance of the hemispheric torquer in an unintentional co-elevation offset.

For the air-bearing model, the rotor marking was determined using a procedure outlined in Appendix B of Ref. 2. The marking was applied by lightly sandblasting one hemisphere and placing the D-pattern on the other hemisphere. The best technique for actually marking the rotor is not obvious; this is an area which needs further exploration. An alternative scheme to completely marking the hemisphere is to duplicate the equatorial pattern with parallel wide and narrow lines (Fig. 16b). Digital logic could then be used to determine whether transitions were into the northern or southern hemisphere. The penalty for such a scheme would be in the area of increased electronics.

3. Optical Sensors

The optical sensors designed and fabricated for the experiment were much larger than needed for an actual gyro. Ease of operation rather than size was the governing factor in their design. Fig. 17 shows a cutaway drawing and a photograph of the sensor. The sensors for all three axes were the same except that the illumination and sensor field stops were larger for the Z-axis sensor. The source of light is an ordinary 6-v bulb that has been painted black. The light passes through a sandblasted circle on the bulb axis. The sandblasting frosts the glass and reduces the possibility of filament vibrations affecting the system. The technique of Koehler illumination [Ref. 4, p. 508] is employed in the optical design to insure uniform sensor response. This technique involves focusing the frosted glass on the objective lens with the illumination lens. In turn, the illumination stop is focused on the rotor using the objective lens. For the reflected light the focus point is the sensor stop after being reflected by the beam splitter. The sensor optics are the dual of the illumination optics in that the image of the photodetector (Fairchild FPM 100 phototransistor) is focused on the objective lens using the sensor field stop lens. The photodetector and preamplifier are attached to the optical sensor in an enclosed housing. There is a provision to include a second photodetector. It would form part of a bridge circuit by receiving a calibrated amount of light directly from the light source through fiber optics. The second detector is used to reduce thermal sensitivities, if this is needed.

One minor design deficiency was the insufficient length of the beam splitter. This caused minor secondary reflections to be received in the photodetector, but fortunately did not prevent satisfactory sensor operation.

4. Electronics

The system electronics were designed and fabricated by the Stanford Inertial Instruments Laboratory. A detailed explanation of the electronics is beyond the scope of this paper. The overall system block diagram is presented in Fig.18. Fig. 19 is a photograph of the electronics and the system control panel. The number of switches and potentiometers was necessary to insure flexibility in the experiment, but they would not be needed for an operational free-rotor gyro. Furthermore, the size of the laboratory electronics is misleading. They were fabricated with discrete components. Using today's packaging techniques and components the additional electronics needed to mechanize hemispheric torquing would be no larger than a cube 1.5 in. on a side.

EXPERIMENTAL RESULTS

1. Hemispheric Torquing (Coarse Damping)

Three basic experiments were performed with the air-bearing model. The first was a measurement of actual damping times when the erection was delayed until $P_2 \leq 0.1$. This experiment provides a check of the analysis presented in Ref. 1. It is not normally the way that active damping should be applied as it results in excessive damping time. The delay simulates the restriction on torquing due to the purely passive spin-axis caging which was used in this version of coarse damping. The formulas for rotor torque are used to find the time

to complete damping in the large. Taking the minimum time after starting to spin-up for the initiation of hemispheric torquing:

$$t_i = \begin{cases} \left[\frac{10 i_{zx}}{i_x}\left(\frac{1}{i_x^2 K_m}\right)\right]^{\frac{1}{2}} & \text{single-axis torquing} \\ \left[\left(\frac{10\sqrt{2}\, i_{zx}}{i_z}\right)\left(\frac{1}{i_x^2 K_m}\right)\right]^{\frac{1}{2}} & \text{two-axis torquing} \end{cases} \quad (4)$$

where i_x = rms current in the X or Y coil

i_{zx} = rms current per phase for hemispheric torquing in the Z coil

K_m = motor constant [Eqs. (1), (2)]

The expected total time for the hemispheric sensors to reach the dentate pattern is derived from Ref. 1.

$$t_t = \begin{cases} t_i \exp\left[\frac{(\pi - \text{sat})}{0.632}\,\frac{i_x}{i_{zx}}\right] & \text{single-axis torquing} \\ t_i \exp\left[\frac{(\pi - \text{sat})}{.9}\,\frac{i_x}{i_{zx}\sqrt{2}}\right] & \text{two-axis torquing} \end{cases} \quad (5)$$

where "sat" is the linear range of the saturation function due to the dentate pattern.

The rotor was spun up to t_i without hemispheric torquing. As nearly as possible this was done 180° away from the desired location in the rotor and approximately caged in the case. Hemispheric torquing was switched on at t_i and the time to reverse the rotor and start to obtain phase markings was recorded. Naturally the rotor took somewhat longer to stabilize completely, but the results could not be validly compared with the results of Ref. 1 due to the effect of the dentate pattern. Fig. 20 shows the comparison of experimental and analytic results. The time ratios t_t/t_i are plotted. It can be seen that the average of the experimental results is in good agreement. Because t_i was in the vicinity of 5 to 10 sec, the variation in results is mainly attributed to the errors in t_i (estimated rms error 0.3 sec). The tendency for the theoretical prediction to be pessimistic for large time ratios is attributed to air-bearing turbine torques slowing down the spin-up. These torques only become sizable for larger time ratios. The larger torques reduce the gyroscopic stiffness of the rotor, thus making the hemispheric torquer more effective.

The second experiment was the measurement of the time to damp in the large when active caging control is used and the hemispheric torquer is turned on simultaneously with spin-up. This approach results in shorter coarse damping times than the approach described in Ref. 1 and is only possible with spin caging control during spin-up. Damping times were checked against a digital simulation of the model which included the effects of amplifier saturation, the saturation function in the caging control, the saturation effect of the dentate pattern, and the effect of turbine torques. Because the rotor was turning very slowly, it was assumed to have a spherical moment-of-inertia ellipsoid in the simulation. The inputs to the simulation were the experimental values of current, motor characteristics, and other parameters of the experiment. Fig. 21 is a plot of the experimental damping times using two-axis torquing for various values of torquing current. Fig. 22 shows the results of the digital simulation for torquing currents $i_x = i_y = 3$ amp and $i_{zx} = i_{zy} = 3$ amp. Four of the experimental run records are shown in Fig. 23. Each is a run for different current levels. The Z sensor in each case starts opposite a point close to the pole of the shiny hemisphere. As damping proceeds, the sensor's locus moves into the dull hemisphere. The final high-frequency oscillations are due to the Z sensor receiving reflections from the shiny D-pattern. The event times from the logic output are also recorded. The four events are defined to be the times at which the rotor has rotated 90°, 180°, 270°, and 360°. One of the events is shown in the lower trace in each record. Each transition level represents an event. Fig. 23 represents samples of a number of runs which were averaged for Fig. 21. Damping was judged complete when the fine damper had the signals necessary for its operation. The damping time for the digital simulation should be compared with the time to start sensing the D-pattern (Fig. 23b).

The experimental results are put in perspective if the total time to reach 1000 rad/sec for a spherical free-rotor gyro with similar rotor characteristics is calculated:

$$t = \frac{1000}{i_x^2\,(0.059)}$$

At $i_x = 4$ amp this is roughly 1000 sec, or about thirty times the active damping time. The reason for the shorter coarse damping times when the damper is turned on immediately is, of course, that it is easier to rotate the rotor before it has built up speed.

The third experiment in the coarse mode performed with the air-bearing model was to allow the gyro to spin up, with the hemispheric torquer on, and stabilize so that there were only small residual motions of the angular velocity vector in the rotor. Then, using a Gaertner M101A microscope mounted on a Nikon cross-traveling micrometer stage, the small motions of the angular velocity vector in the rotor were measured, using the method developed in Ref. 5. The maximum oscillatory amplitude in the pointing direction of the angular velocity vector was 100 ± 20 arc sec. The hemispheric torquer was then turned off and the rotor was allowed to polhode. The maximum polhode amplitude observed was 140 ± 20 arc sec. This experiment was conducted with the rotor spinning at 2500 rpm. This shows that hemispheric torquing may have the potential of adequately damping the polhode motion with no need for fine damping.

2. D-Pattern Damping (Fine Mode)

The fine damping using the D-pattern was also performed on the air-bearing model. The experimental sequence follows:

1. With rotor upside down as a worst-case initial condition, coarse damping is initiated simultaneously with spin-up (active caging is also used).

2. When the event times are occurring regularly (about 30 sec after start), the automatic delayed switch is started to change from coarse to fine damping (the delay is roughly 2 min).

3. After the system switches into fine damping, spin-up is continued until a total of 12 min. has elapsed. At this time, the completion of the erection cycle is simulated by turning off the D-pattern polhode damper.

4. The polhode amplitude is experimentally determined after system shutdown; this is a measure of the quality of the initial alignment.

The primary expected sources of error were the "hangoff" (a torque switching at rotor frequency after the polhode is damped) of the polhode damper and the additive noise in the D-pattern sensor. Because the free-rotor gyro is being modeled with the air bearing, a number of additional undesired effects probably are present which would not be observed with an ESG or with an Unsupported Gyro:

1. Geometry-dependent turbine torques which have body-fixed components. These may be functions of air-bearing gas pressure, temperature, and the hemisphere to which the spin-axis is being erected.

2. As a result of earth rate and inertially fixed turbine torques, automatic spin-caging torques are quite large. Although the error signal is passed through a lowpass filter, it still may have body-fixed components.

3. Although the rotor is made of beryllium-copper to avoid permanent magnetism, a small amount of permanent magnetism has been measured on the rotor (0.6 mgauss). During operation, the magnetic dipole may move in the rotor leading to undesirable motion of the apparent maximum axis of inertia.

In spite of these effects, the fine damping consistently reduced the amplitude of polhode motion to 5 arc sec. Since the purpose of removing this polhoding is to eliminate perturbations on the spin-axis readout, the natural question is to ask what the effect of 5 arc sec of polhode amplitude would be on a typical free-rotor gyro readout system. Since the polhode and spin frequencies typically differ by a factor of 10 or greater, the readout error may be estimated to be an order of magnitude smaller (1/2 arc sec) or less. Furthermore, it should be pointed out that the polhode error magnitude due to hangoff is very nearly

$$e = \frac{M_h}{I_z \beta^2 \sqrt{\epsilon_1 \epsilon_2}}$$

where M_h = undesired disturbance torque

I_z = Z-axis moment-of-inertia

β = spin speed (rad/sec)

ϵ_1, ϵ_2 = moment-of-inertia difference ratios

If it is assumed that the disturbance torque, including the additional effects mentioned above, is primarily due to hangoff and that the experimental results are of a typical magnitude, then the effect on a hollow, free-rotor gyro may be estimated by considering that

1. I_z of a hollow rotor is roughly one-tenth that of the model.

2. β is 3 to 10 times that of the model (model spins at 3000 rpm).

3. ϵ is greater by about 100 (the model was 3.1×10^{-4}).

The conclusion that the model's residual error may be expected to be reduced by a factor of 100 or more in an actual ESG is probably false due to reasons explained in the next paragraph. The point is that the effect of hangoff of the same magnitude will be less in most free-rotor gyros.

The primary reason that so great a reduction would not be realized is sensor output disturbance. If the main reason for the residual polhode is sensor noise, which will be assumed to be white and Gaussian, then the hangoff is

$$\text{rms } (e) = \left(\frac{3 N_o K_p}{2 \beta \pi} \right)^{\frac{1}{2}}$$

where K_p = polhode damper loop gain (cf. p.90 Ref.2)

N_o = noise power per cycle in the sensor output

β = spin speed.

Typical experimental recordings are presented in Fig. 24. The upper trace shows the output of the Z sensor during the final stages of damping; that is, the polhode damper is still on. The lower trace has a coarser scale and shows the residual polhode of a typical run just after turning off the polhode damper.

The statistics of the experimental results are as follows:

Number of runs	26
Mean residual polhode amplitude	5.6 arc sec
Experimental standard deviation	2.0 arc sec

ELIMINATION OF FINE DAMPING

Some experimental work has been performed on eliminating fine damping and using coarse damping alone. This section will offer the reasons why such an approach may be possible, together with the very sketchy experimental evidence that supports this possibility.

Based on experimental observations, hemispheric torquing and active caging will damp the polhode motion below levels that can be measured from the Z-axis output. The rotor-fixed equilibrium position is not the maximum axis because of slight mismarking of the rotor and the consequent body-fixed torques. If, after damping the polhode motion, the hemispheric torquer is turned off, the system will begin to polhode about a new equilibrium point. This point is close to the maximum axis of inertia but not exactly aligned because of those effects discussed in the last section which are not due to the damping electronics. This motion, for all practical purposes, is undamped.

To eliminate fine damping, there are two possible approaches. The first approach is simply to apply an additional torque, oscillating at rotor frequency, which has been carefully calibrated to cancel the body-fixed error torques caused by the hemispheric torquer. The advantage of this method is that it is relatively simple to implement and calibrate. The disadvantage is that the apparent error may be slowly changing, so the validity of the calibration may be questionable.

The second method is illustrated in Fig. 25. Rough experimental checks tend to establish the body-fixed perturbing torque (due to the polhode loop) as having a fixed direction with magnitude proportional to the hemispheric-torquing currents. If at some time T = 0, the hemispheric torquer is only partially shut down and then one-half polhode period later is completely shut down, the motion of the spin axis in the rotor would be as shown in Fig. 25. This technique will be called "stutter shutdown". If the model for the perturbing torque is correct, this method would work in spite of day-to-day changes in the effective torque error. At first thought, it would seem that the spin speed would affect the time of shutdown, but the number of rotor cycles per half period of polhode motion is an invariant quantity for low damping ratios. The number of rotor cycles is

$$\text{Rotor cycles} = \frac{1}{2\sqrt{\epsilon_1 \epsilon_2}}$$

The disadvantage of stutter shutdown is that the initial calibration may be somewhat difficult. Very sketchy initial experimental results indicate that polhode levels of less than 20 arc sec are possible.

Both of these techniques should be considered in practical uses of the active damper because they help achieve a principal engineering goal of system simplicity.

CONCLUSION

1. General

The use of active damping to move the spin-axis in a freely spinning spherical gyro rotor to a desired location has been shown to be both feasible and practical. A control law to accomplish the damping has been shown. The advantage, in damping time, over the methods of passive damping now used is very conservatively estimated to be an order of magnitude. The scaling of the problem is such that damping should always be completed by the time spin-up is complete. This is, a gyro that can be spun up in 2 to 3 min. can be damped in the same time. Complete experimental verification of the theoretical results was obtained using an air-bearing model, and the slight differences between the theoretical results of Refs. 1 and 2 and the experimental results reported here can be attributed to two main effects: turbine torques developed by the air bearing, and uncertainties in the initial conditions. To use this method with a current ESG, no new sensors or coils are required. The increased volume of electronics has been estimated to be a cube 1.5 in. on a side, with state-of-the-art techniques. This quantity of electronics could possibly be further reduced in a complete guidance by including a gyro spin-up mode in the guidance computer. For use in an ESG, the only major remaining question is how best to mark the rotor.

2. Hemispheric Torquing

The method of hemispheric torquing was demonstrated analytically and experimentally. The method can be used in either a single or a two-axis mode, but comparison with the time optimum for fixed erection-torque magnitude shows the latter to be the more effective by factors of up to 2. For use with passive caging, the rotor should have some initial angular velocity in order to preclude instabilities. When this condition is met, an accurate formula (verified experimentally) for two-axis torquing is

$$t_t = t_i \exp\left[\frac{(\pi - \text{sat})}{0.9} \quad \frac{i_x}{i_{zx}\sqrt{2}}\right] \qquad (5)$$

where t_t = time to damp the spin-axis to the proportional range of the dentate pattern

t_i = time, after spin-up initiation, at which the coarse damper is turned on

i_x = current in each spin-up coil

i_{zx} = current per phase in the control coil

This result shows that damping times of less than 25 percent of the time to spin up may be easily attained in typical spherical free-rotor gyros. A dentate rotor pattern was developed that allowed a saturation type of damping control for the hemispheric torquer. It has a characteristic that is superficially similar to a pattern developed at the University of Illinois for ESG readout [reported in Ref. 6], but the use and readout of the dentate pattern are quite different.

Using the dentate pattern, and with very little added electronic complexity (and no additional coils or sensors, if D-pattern spin-axis readout is used), an active caging system (in the large) can be added. Active caging leads to faster damping because the hemispheric torquing may be initiated simultaneously with rotor spin-up.

The steady-state offset of the hemispheric torquer was determined by turning it off and allowing the rotor to polhode. The motion, as observed by a Gaertner M101A microscope mounted in a cross-traveling micrometer stage, had an amplitude of 140 arc sec. More precision in the marking scheme should allow this amplitude to be reduced by a factor of 2 or 3, which should be sufficient for many applications without recourse to a fine-damping system. A method of shutting down the hemispheric torquer, called stutter shutdown, offers promise of even smaller residual polhode amplitudes without recourse to fine damping.

3. Fine Damping with the D-Pattern

In addition to the experimental work reported here, the control law for fine damping using the D-pattern was also validated with an analog simulation (Ref. 2). Included in the simulation were the effects of sensor saturation and controller null offset. Sensor saturation reduced the speed of damping but led to no steady-state errors. The controller null offset, on the other hand, showed polhode offsets of 7 arc sec for a current 1/1000 of maximum which was switching in synchronization with the rotor. The formula for this offset is

$$\left(\frac{W_y}{\beta}\right)_{ss} = \frac{M_h}{\beta^2 \epsilon} \quad \text{(radians)}$$

where M_h is the hangoff torque, β is the rotor spin speed, and ϵ is the moment-of-inertia difference ratio.

This shows that increasing rotor speed β will decrease the effect and that rotors with larger moment-of-inertia difference ratios will also be less sensitive to this problem. The effect of sensor output disturbances will not be reduced as much by higher rotor spin speeds and larger difference ratios. For Gaussian white noise with power spectrum of N_o, the rms offset is

$$\langle (W_y/\beta)^2 \rangle^{1/2} \propto (N_o/\beta)^{1/2}$$

The D-pattern damper should be capable of reducing polhode motion to less than 5 arc sec in practice.

The theoretical results were confirmed by the experimental model. The average polhode amplitude after the damper was turned off was 5.6 arc sec.

4. Future Work

a. The active damper should be incorporated in an existing ESG, to find out if there is some incompatibility that is not presently apparent.

b. Further investigation of the methods of marking the gyro rotor should be made. The goal would be to minimize the effect of markings on the inertia, mass unbalance, and dielectric properties. This is discussed in more detail below.

c. For caging in the large, using the dentate pattern, the possibility of using a fixed time delay, rather than the Z-axis sensor, to switch the sign of caging gain should be investigated. A sketchy experimental check suggests this is possible, but further investigation of possible instabilities will have to be made before it can be accepted.

d. Further investigation of low-frequency, optical-sensor output disturbances should be undertaken to determine the sources of error and potential accuracy of the D-pattern damper.

e. The steady-state hemispheric-torquer error (140 arc sec) is better than the basic marking accuracy by a factor of about 3. This may be a random result, which, since there is only one sample, cannot be confirmed at this time. There may be another reason or reasons. The possibility should be explored that the dentate pattern leads to some nonlinear form of polhode damping, or that the combination of (a) constraining a rotor to a given rotor-fixed spin-axis (not the maximum axis of inertia) and (b) metallic creep tends to make the maximum axis coincide with the constrained axis.

f. A detailed, quantitative investigation of the stutter shutdown technique should be performed. The desired results would include the determination of the exact shape of the spiral and measurement of the equilibrium point vs. hemispheric-torquing amplitude.

5. Application to Existing Electrically Supported Gyroscopes

The rotor markings used in these experimental studies were developed with two goals in mind:

a) Simplicity of construction for the laboratory electronics to verify the basic principle of active damping, and

b) Possible application in an unsupported gyroscope (Ref. 7) where (because the rotor is in free fall in space) the markings do not appreciably affect the drift rates.

In order to apply this technique to conventional ESG's additional work is necessary. There are other methods by which the hemispheres may be coded than by making the entire surface dull. Fig. 16b shows one possible method, and there exist other techniques for hemispheric coding. One very interesting point concerning the need for active damping is that with active damping it is always possible to choose which hemisphere contains the "north" pole. With magnetic or structural damping the rotor may damp in with either side up. Thus the user of ESG's without active damping in a navigation system has the choice of providing two error models per gyro or of engaging in multiple start-up attempts until all gyros are properly aligned. Thus it can be seen that the active damper offers a considerably greater time saving than would be predicted merely by considering the ratio of damping time with a magnetic field to active damping time. In this connection if the two hemispheres are canted on the rotor so that their dividing plane makes an angle approximately 45° with respect to the rotor maximum axis of inertia, preliminary experimental results indicate that the hemispheric torquer will still tend to drive the spin-axis to that axis of maximum inertia in the rotor which is closest to the normal to the dividing plane if the hemispheric torquer moments are not too large. Thus in a strap-down ESG system which uses a great-circle cosine-readout system and which codes the hemispheres, hemispheric torquing could be used in conjunction with magnetic field damping to always guarantee that correct side was up and to hasten the damping time.

Finally, it should be remarked that the fine damper was included in this study primarily to find out how well one could do by these techniques and because it would be needed for an Unsupported Gyroscope (Ref. 7). In an operational ESG system the primary advantage of active damping, as mentioned above, is that the gyro always damps in with the same side up. Thus it would probably be best to omit fine damping in an operational system and depend on either calibrating out the gyro marking error, using stutter shutdown, reducing the hemispheric torquing moments at the end of the damping time so that the spin-axis is driven toward the maximum axis more than toward the effective pole of the markings, or using magnetic damping in connection with hemispheric torquing.

It seems almost certain that some combination of the above techniques will remove the need for fine damping in operational systems, and this will greatly reduce the complexity of the associated electronics. This is true because a sizable fraction of the electronics used in this experiment is already included in an ESG for initial spin-up and for spin-axis readout.

REFERENCES

1. Parkinson, Bradford and Lange, Benjamin, "The Active Damping of Free-Rotor Gyroscopes During Initial Spin-up", Proceedings of the AIAA/JACC Guidance and Control Conference, Seattle, Washington, August 15-17, 1966, pp. 673-687.

2. Parkinson, Bradford and Lange, Benjamin, "The Active Damping of Free-Rotor Gyros", Stanford University Department of Aeronautics and Astronautics Guidance and Control Laboratory SUDAAR Report No. 260, May 1966.

3. Skilling, H. H., Electromechanics, John Wiley and Sons, Inc., New York, 1962.

4. Hardy, A. C. and Perrin, F. H., The Principles of Optics, McGraw-Hill Book Company, New York, 1932.

5. Plank, R. V., "Optical Flat Placement on Almost Spherical Gyro Rotors", Engineer's Thesis, Department of Aeronautics and Astronautics, Stanford University, August, 1965.

6. Baker, D. H. and Harrill, J. W., "Basic Principles of Unconventional Gyros", M.S. Thesis, Massachusetts Institute of Technology, Mary 1954.

7. Lange, Benjamin, "The Unsupported Gyroscope", Proceedings of the 1964 Unconventional Inertial Sensors Symposium, (Proceedings are classified, but the paper is unclassified), New York, November 1964.

FIG. 1. AIR-BEARING MODEL OF THE FREE-ROTOR GYRO. (The rotor diameter is 2 in. Other sizes may be judged by the 6-in. ruler on the base plate.) The three optical sensors are located on the upper three faces of the cube which constitutes the gyro case.

FIG. 2. AIR-BEARING GYRO ROTOR. The solid beryllium-copper rotor has been precisely machined to a 2-in. diameter. Sphericity is better than 25 μin. Clearly visible are the D-pattern, the dentate equatorial pattern, and the marking modification of the dentate pattern to enable rotor phase to be determined. Note that the double-tooth phase marking lines up with the edge of the D.

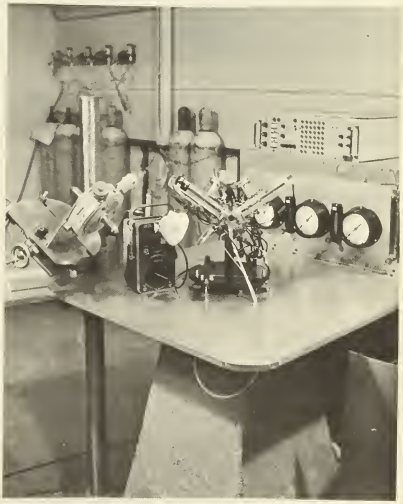


FIG. 3. AIR-BEARING GYRO ON STAND. The pressure and flowmeters used to monitor the air-bearing behavior may be seen at the rear of the stand. To the left is the microscope and cross-traveling micrometer stage. The nitrogen supply is visible in the background.

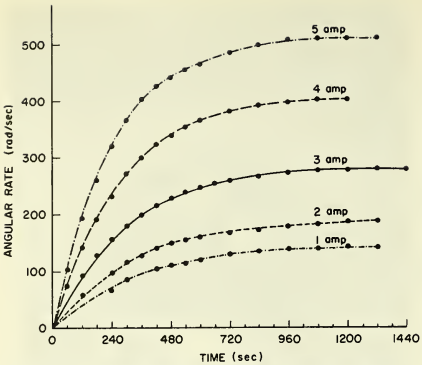


FIG. 4. AIR-BEARING GYRO MOTOR, STEP-RESPONSE CURVES FOR VARIOUS RMS CURRENT LEVELS IN THE X AND Y COILS. Data were taken with a stroboscope and stopwatch.

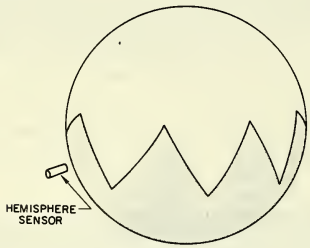


FIG. 5. DENTATE ROTOR PATTERN

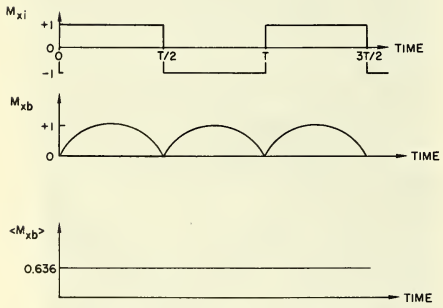


FIG. 6. INERTIAL AND BODY-FIXED TORQUES FOR LARGE COELEVATION ANGLES USING SINGLE-AXIS HEMISPHERIC TORQUING. The rotor spin period is T. M_{xi} is the torque applied in the inertial reference frame (along the X axis); M_{xb} is the projection of the torque onto the rotor X axis; and M_{xb} is the average torque in the rotor coordinate system.

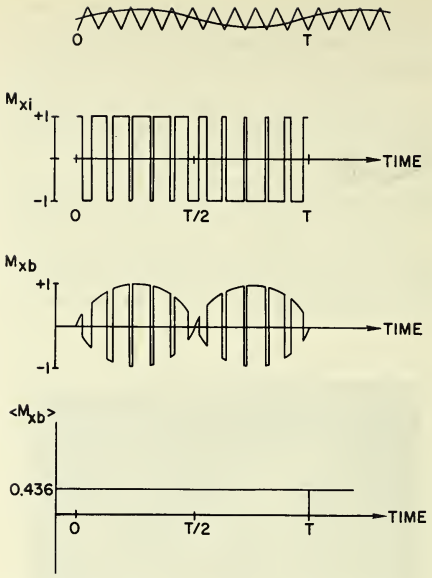


FIG. 7. INERTIAL AND BODY-FIXED TORQUES FOR SMALL COELEVATION ANGLES. This diagram shows the reduction in average rotor-fixed torque as the sensor locus no longer clears points of the teeth. The upper picture shows the locus of the sensor through the dentate pattern. The notation is that of Fig. 6.

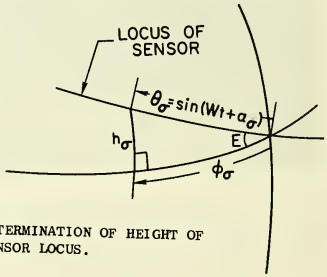


FIG. 8. DETERMINATION OF HEIGHT OF POINT ON SENSOR LOCUS.

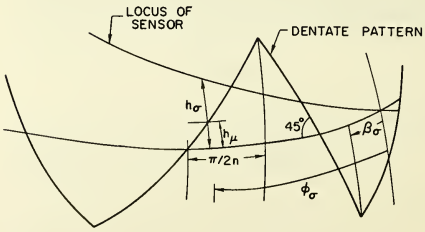


FIG. 9. DETERMINATION OF TOOTH HEIGHT OF DENTATE PATTERN.

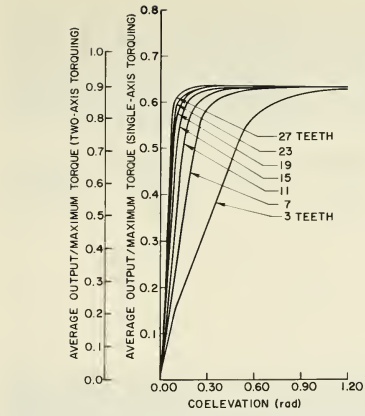


FIG. 10. HEMISPHERIC-TORQUING SATURATION CHARACTERISTIC DUE TO THE DENTATE PATTERN.

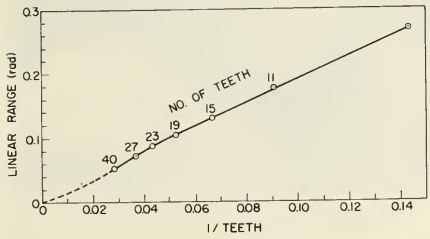


FIG. 11. LINEAR RANGE OF HEMISPHERIC TORQUER vs. INVERSE TEETH. The tooth edges are great circles crossing the equator at 45°. Flatter teeth will decrease the linear range for the same number of teeth.

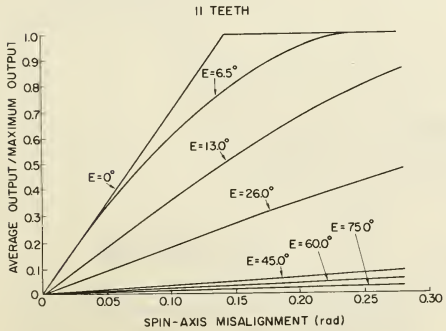


FIG. 12. SPIN-AXIS MISALIGNMENT SENSITIVITIES FOR VARIOUS COELEVATION ANGLES (E) USING AN 11-TOOTH DENTATE PATTERN.

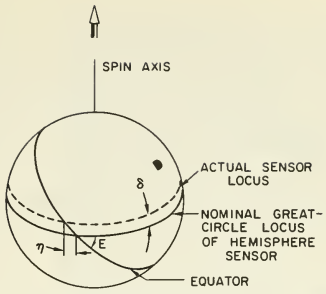
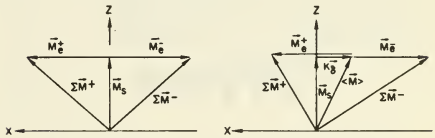


FIG. 13. THE EFFECT OF SPIN-AXIS MISALIGNMENT FOR LARGE COELEVATION ANGLES. "Large" may be taken as coelevation angles (E) larger than 45° for the 11-tooth pattern.



$\vec{M}_e^+$ = POSITIVE ERECTION TORQUE

$\vec{M}_e^-$ = NEGATIVE ERECTION TORQUE

$\vec{M}_s$ = SPINUP TORQUE

$K\vec{\delta}$ = SPIN CAGING TORQUE

$\langle\vec{M}\rangle$ = AVERAGE TORQUE

$\Sigma\vec{M}$ = TOTAL TORQUE

FIG. 14. TOTAL TORQUE IN XZ PLANE WITH AND WITHOUT SPIN-CAGING CONTROL.

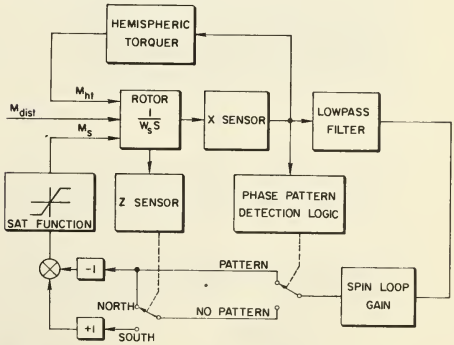


FIG. 15. BLOCK DIAGRAM OF X-AXIS SPIN-CAGING CONTROL. The Y axis is similar. In the experiment the spin loop gain is set fairly high to reduce the effect of air-bearing turbine torques. The logical decisions illustrated are for active damping whose goal is the northern hemisphere along the positive Z case axis. M_{dist} is a disturbing torque and M_{HT} is the control torque.

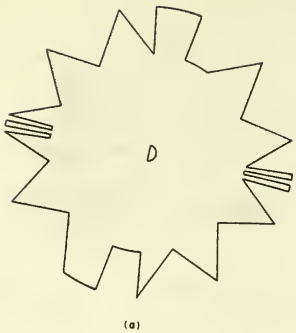


FIG. 16. DENTATE PATTERN CHARACTERISTICS. Pattern (a) is that actually on the 11-tooth model. Pattern (b) shows only a portion of the equator and is an alternative to completely marking a hemisphere. The dentate pattern is marked in parallel narrow and broad lines. The hemisphere is determined logically from whether a broad or narrow line was last sensed.

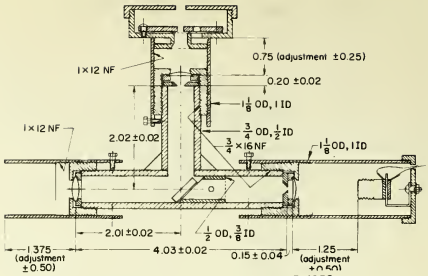
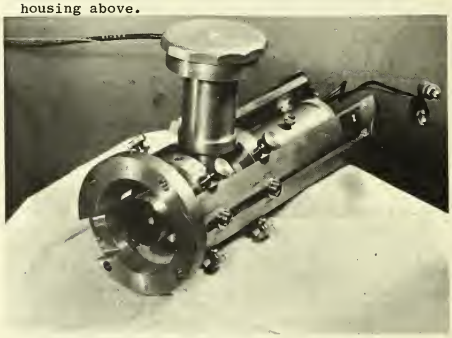


FIG. 17. DETAILS OF THE OPTICAL SENSOR.

a. Cutaway drawing. The rotor would be to the left, the light source in the socket to the right, and the preamplifier enclosed in the housing above.



b. Photograph showing mounting bracket and position-adjusting mechanism.

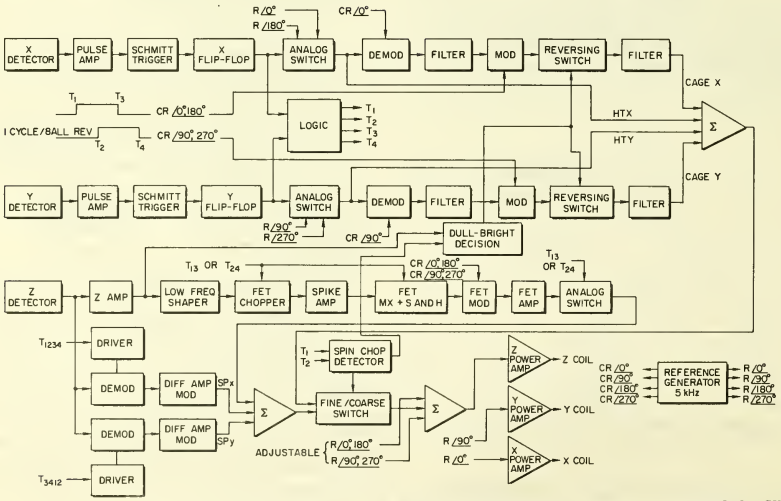


FIG. 18. BLOCK DIAGRAM OF THE SYSTEM ELECTRONICS FOR THE EXPERIMENTAL AIR-BEARING, FREE-ROTOR GYRO.

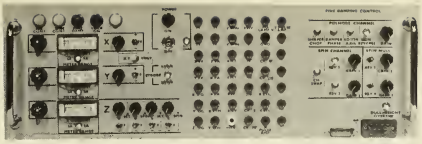


FIG 19. ELECTRONICS PANEL SHOWING THE VARIOUS CONTROLS AND GAIN POTENTIOMETERS NEEDED FOR EXPERIMENTATION.

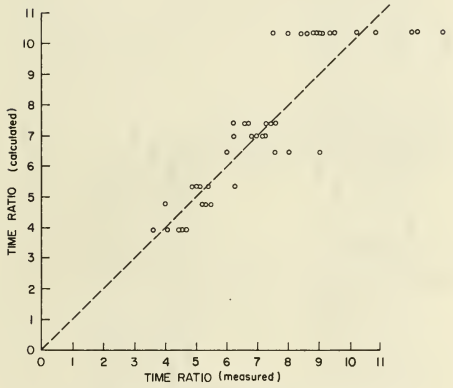


FIG. 20. EXPERIMENTAL RESULTS WITH SIMULATED, PASSIVE CAGING. For these data the gyro was first allowed to attain a nominal amount of spin, then the active damper was switched on. The ratio of damping time to start time is a constant (analytically) for a given ratio of hemispheric control torque to spinup torque. For exact agreement, the results would have "plotted" along the 45° line.

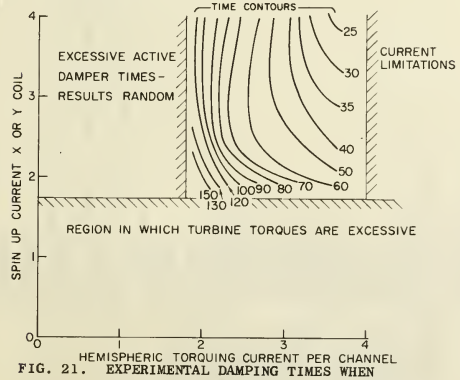


FIG. 21. EXPERIMENTAL DAMPING TIMES WHEN DAMPING SIMULTANEOUSLY WITH INITIAL SPINUP. All times are in seconds. Initial condition is the rotor upside down in the case. The spinup motor delivers a torque of 0.059 i_x^2 rad/sec^2.

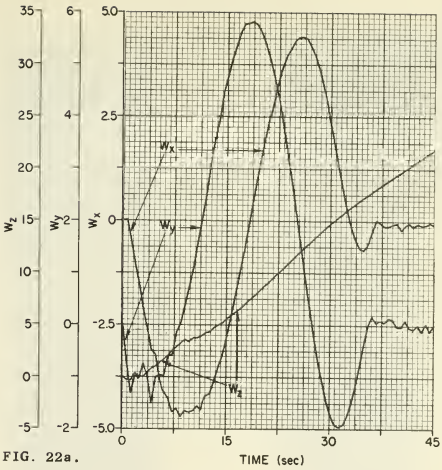


FIG. 22a.

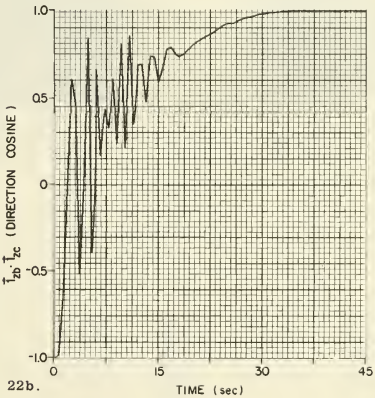


FIG. 22b.

FIG. 22. RESULTS OF DIGITAL SIMULATION OF THE AIR-BEARING MODEL. $E_o = \pi$, $i_x = i_y = 3.0$ amp. The direction cosine starts out at -1.0 which indicates that the rotor is upside down. W_x, W_y, and W_z are the three body rates, measured in the rotor. The desired final condition is $W_x = 0$, $W_y = 0$.

67-590

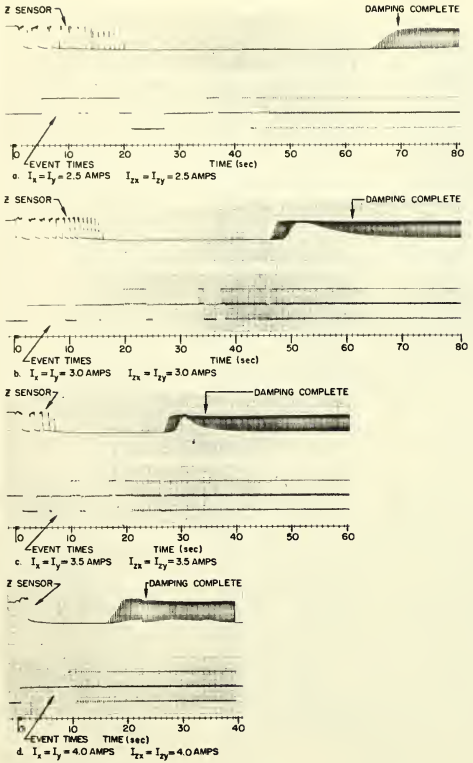


FIG. 23. EXPERIMENTAL RUN RECORDS FOR ACTIVE CAGING. In each case the Z sensor first sees the wrong (bright) hemisphere, then sees the dull hemisphere, and finally senses the shiny D on the dull hemisphere. Damping in the large is complete when the steady D is being sensed and several seconds of the event times have been present, so the system may be shifted to fine damping.

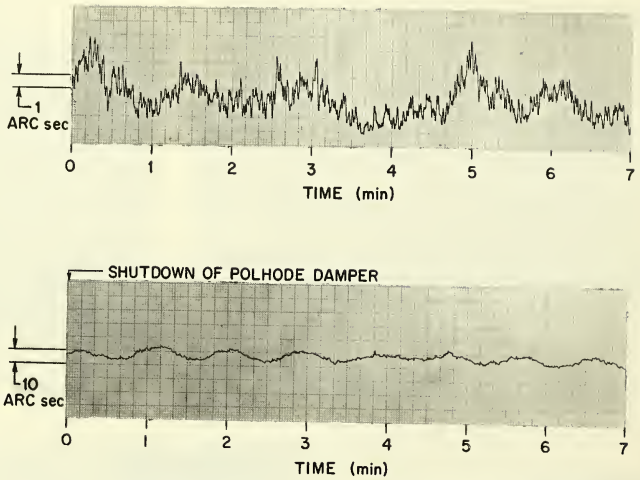


FIG. 24. EXPERIMENTAL RESULTS OF D-PATTERN POLHODE DAMPING. The upper trace is the D-pattern sensor output with the controller on. The lower trace shows the residual polhode after the control is shut off. The rotor had a marking error of about 300 arc sec; the lower trace shows the ability of the polhode damper to damp to the maximum axis of inertia in spite of marking errors.

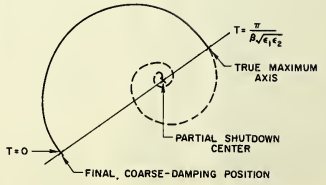


FIG. 25. THE MOTION OF THE SPIN AXIS IN THE ROTOR DURING STUTTER SHUTDOWN. Coarse damping is partially shut down at T = 0; one-half polhode period later (when the spin axis is very close to the maximum axis of inertia) coarse damping is completely shut down.

No. 67-591

ADVANCED CONTROL SYSTEMS FOR THE SATURN V LAUNCH VEHICLE

by

MICHAEL T. BORELLI and STANLEY N. CARROLL
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-591

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.02, 2.01, 3.03

-- NOTES --

ADVANCED CONTROL SYSTEMS FOR THE SATURN V LAUNCH VEHICLE

Michael T. Borelli and Stanley N. Carroll
Flight Dynamics Branch, Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

Four techniques of adaptive control have been studied for possible use in stabilizing the structural bending of large launch vehicles. The program was initiated to develop a backup control system for the Saturn V vehicle. The four control system configurations are a tracking notch filter, a tracking offset notch filter, a phase-sensing system, and a type of model reference adaptive system. Considering overall vehicle performance, the tracking offset notch filter system, which provides phase stabilization for the structural bending mode, was shown to be superior to the other three systems. A miniaturized single-axis prototype of the offset notch filter configuration has been built and successfully tested in an analog computer simulation.

Nomenclature

ϕ_c	attitude command
$\phi_{8.23}$	attitude angle measured at 8.23 calibers from the engine gimbal plane
$\dot{\phi}_{4.0}$	attitude angle rate measured at 4.0 calibers from the engine gimbal plane
$\ddot{r}_{7.0}$	lateral acceleration measured at 7.0 calibers from the engine gimbal plane
β, β_c	engine deflection angle and engine command angle, respectively
ϵ	filtered bending mode signal
K_D, K_R	attitude gain and attitude rate gain, respectively
K_A, K_{AD}	acceleration gain and adaptive loop gain, respectively
K_{N1}, K_{N2}	gains associated with tracking bandpass filters
P_n	square wave commutation functions ($n = 1,2,3,4$)
s	Laplace variable ($\sigma + j\omega$)
$Y_{i(\bar{x})}$	bending mode deflection of ith mode at station $\bar{x}$
$Y'_{i(\bar{x})}$	bending mode slope of ith mode at station $\bar{x}$

Introduction

The development of large launch vehicles has greatly intensified the structural bending stabilization problem. Three of the main factors contributing to this control problem are (1) the proximity of the vehicle control frequency to the first structural bending mode resonant frequency; (2) the uncertainty associated with determining the structural bending mode parameters, both theoretically and experimentally; and (3) the extremely low structural damping of the bending modes. During the definitive stage in the development of a vehicle such as Saturn V, the possibility that conventional control systems will not adequately handle all these problems must be considered. This becomes increasingly important if the proposed vehicle is to perform a wide range of missions using a variety of payloads. To nullify the impact of such an eventuality, MSFC's Astrionics Laboratory in conjunction with the Martin Company of Denver, Colorado, has been investigating four types of advanced control systems for the Saturn V vehicle.

The most promising system resulting from this study effort is an adaptive tracking notch filter that has its notch offset a prescribed amount from the bending mode frequency. The intentional offset in frequency of the notch provides phase lead at the particular bending mode frequency and thereby achieves the desired stabilization. This configuration is referred to as an adaptive tracking offset notch (ATON) filter.[1,2]

The other three systems investigated were a tracking notch filter for gain stabilization, a phase-sensing system, and a type of model reference system. In both the phase-sensing and model reference systems, an adaptive gain adjustment regulates the gain in a feedback path to acquire the desired stabilizing signal.

For this study, the tracking filters were mechanized using commutated networks.[3,4,5,6] The frequency identification was accomplished by a phase-locked oscillator. Other means for mechanizing the filters and achieving frequency identification could be employed with similar results as far as vehicle performance and stability studies are concerned.

A major objective of this study was to develop a control system capable of tolerating large variations in bending mode parameters, notably the bending frequency and the bending mode slopes (or bending signal gain). No attempt was made to provide any type of adaptive control in the rigid-body frequency interval. A criterion was established to maintain a minimum of 6 db gain margin and 30 degrees phase margin for the rigid-body.

Adaptive Tracking Notch Filter System

The adaptive tracking notch (ATN) filter system is shown in Figure 1. The subscripts on the airframe variables denote the distance in calibers of the sensor locations from the engine gimbal plane.

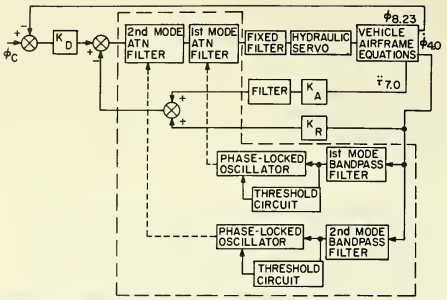


FIGURE 1. ATN FILTER CONTROL SYSTEM.

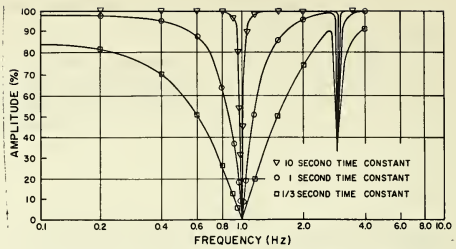


FIGURE 3A. ATN FREQUENCY RESPONSE, AMPLITUDE.

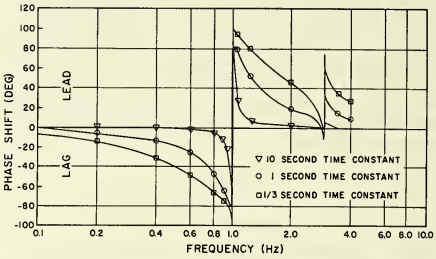


FIGURE 3B. ATN FREQUENCY RESPONSE, PHASE

In this configuration, notch filters are employed to gain stabilize both the first and second bending mode signals. The maximum attenuation of each filter is about 40 db. The filters were designed to track over approximately one octave; the first bending mode tracker range is from 0.7 to 1.4 Hz, and the second bending mode range is from 1.5 to 3.0 Hz. The frequency identification circuit is separate from the control loop; its purpose is to identify the particular bending mode frequency and to adjust the notch filter center frequency to coincide with the bending mode frequency.

The notch filter is shown in Figure 2. The frequency response characteristics of the notch filter are dependent

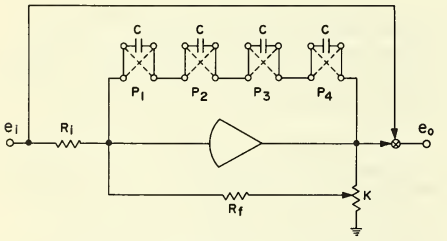


FIGURE 2. NOTCH FILTER NETWORK.

on the RC time constant as shown by the family of curves in Figures 3A and 3B. To use the tracking notch filter for gain stabilizing a bending mode, one must select a time constant that will provide sufficient attenuation not only at the center frequency, but also in the proximity of the center frequency. It is readily obvious that the tracking accuracy becomes increasingly important in this regard since the accuracy in positioning the notch will influence the choice of the time constant. The penalty that results from using a filter with a wide notch (small time constant) is the additional phase lag introduced in the rigid-body frequency region.

Once the time constant has been selected for the expected variation in bending mode slopes, the system has no capability to adapt to any additional variations that may occur in the bending mode slopes. Thus, in addition to considering tracking accuracy, the filter's time constant must also be selected on the basis of the expected slope variation plus a realistic tolerance. This generally results in a relatively broad notch filter for the first bending mode. Also, if the actual bending frequency is below the theoretical values, the adaptive first mode filter will cause additional phase lag at the rigid-body frequency because the filter will track down from its initial position. This occurrence is highly undesirable because the rigid-body phase margin is generally smaller than desired.

In summary, the tracking notch filter is adaptive to bending frequency variations but is non-adaptive to bending mode slope (gain) variations. The severity of the tradeoff between the width of the first mode notch filter and the rigid-body control margins is directly dependent upon how large a total variation is expected in the first bending mode frequency and slope and the accuracy of frequency identification.

Phase-Sensing and Modified Model Reference Systems

The phase-sensing and the modified model reference systems[1] have several features in common. The phase-sensing system, shown in Figure 4, contains a conventional attitude, attitude rate, and accelerometer feedback loop, plus an additional feedback signal through a bandpass filter which is multiplied by an adaptive gain, K_{AD}.

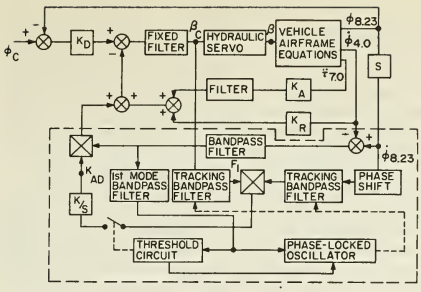


FIGURE 4. PHASE-SENSING ADAPTIVE CONTROL SYSTEM.

For this system, K_{AD} is adjusted so that at the closed loop bending pole frequency a 90-degree phase relationship will exist between the attitude signal resulting from bending and the engine gimbal signal. The modified model reference system, shown in Figure 5, is basically the same as the phase-sensing system; the major difference between the two is the adjustment of K_{AD}. In the modified model reference system, K_{AD} is adjusted so that the closed loop bending poles will be at a specific frequency. It was concluded from the stability studies that the phase-sensing system is somewhat superior to the modified model reference system. Both systems were designed to adaptively adjust the variable gain to compensate for variations in the bending mode parameters. However, neither system proved adequate to handle variations in the bending mode frequencies, although both maintained vehicle stability for very large variations in the bending mode slopes.

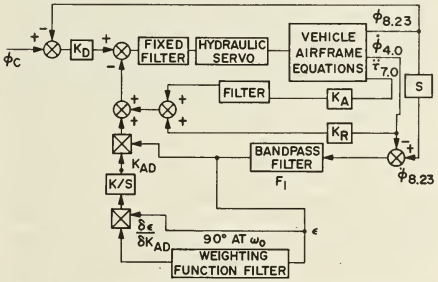


FIGURE 5. MODIFIED MODEL REFERENCE ADAPTIVE CONTROL SYSTEM.

In the modified model reference system, root locus plots were made for the nominal and off-nominal conditions where K_{AD} was treated as the variable. In Figure 6A, loci are plotted for the nominal bending mode frequency and a ±20 percent variation from nominal. Two loci are shown for the +20 percent tolerance case; the upper locus is the bending pole locus and the lower locus results from a pole introduced by the filter F_1. Because the loci are relatively close together, both must be considered for stability studies. For this particular tolerance case, the system operating frequency is established by the filter pole locus and not the bending

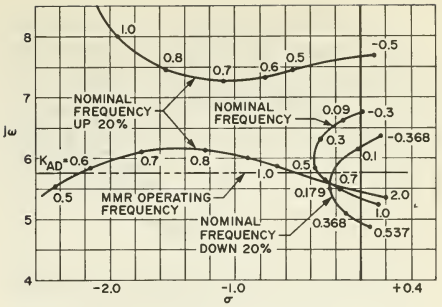


FIGURE 6A. ROOT LOCUS FOR MODIFIED MODEL REFERENCE (MMR) SYSTEM, MODE FREQUENCY VARIATION.

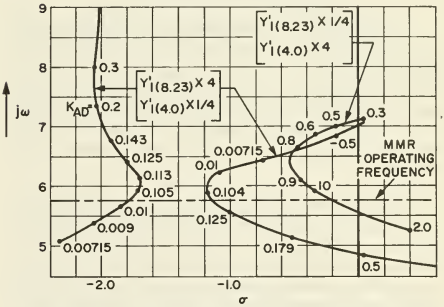


FIGURE 6B. ROOT LOCUS FOR MODIFIED MODEL REFERENCE (MMR) SYSTEM, MODE SLOPE VARIATION.

pole locus. For the nominal and -20 percent tolerance cases, the filter pole loci are not in the proximity of the bending pole loci. Loci were also obtained for combinations of slope variations at the rate gyro locations (Fig. 6B). Dual loci in close proximity again occur for one of the mode slope tolerance cases. However, the filter pole locus, located on the far left, does not present any significant problems in the operation of the adaptive system. From the root locus plots, a closed loop operating frequency of 5.8 rad/s was selected. Each locus must pass through the selected frequency in the left half plane to ensure stability. Two undesirable characteristics associated with this system are the reduced stability margins for the off-nominal frequency cases and the destabilizing influence of the first mode adaptive gain on the second mode bending poles. Both of these effects result from the frequency characteristics of the bandpass filter, F_1. The pass band of the filter is determined by the bending frequency at liftoff and at burnout. The purpose of the filter is to prohibit K_{AD} from affecting signals other than the bending mode signal being stabilized. Making the filter highly selective to accomplish this purpose has the adverse effect of reducing the first bending mode margins for the off-nominal frequency conditions because of the rapidly changing phase characteristics associated with a highly selective filter. The gain-phase plots of Figure 7 show the foregoing effect for the nominal bending frequency and the -20 percent variations from nominal.

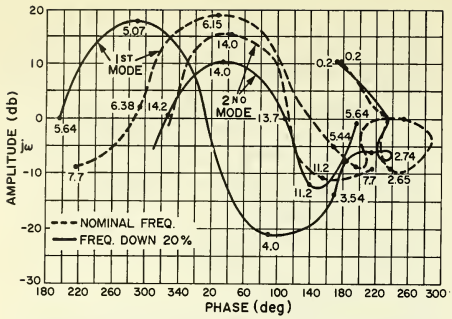


FIGURE 7. OPEN-LOOP FREQUENCY RESPONSE AT OPERATING K_{AD} FOR MODIFIED MODEL REFERENCE SYSTEM.

In stabilizing both first and second bending modes with the modified model reference control system, a very significant problem existed in eliminating the coupling effects and at the same time maintaining adequate margins for the off-nominal conditions. No immediate solution to this problem was apparent; thus the model reference idea was dropped in preference to the phase-sensing system.

For the phase-sensing system, the objective is to maintain a fixed phase relationship between two signals. The manner in which this system functions is not as obvious as that of the modified model reference system because of the difficulty in ascertaining the value of the closed loop bending frequency. For this system, it was found that the adaptive gain adjustment will drive the closed loop bending mode frequency approximately equal to the open loop bending mode frequency.

A vector plot representing the open loop gain through each of the feedback paths is used to explain the phase-sensing system functions. A typical plot is shown in Figure 8. The vectors representing the open loop gain through the attitude, attitude rate, and accelerometer loops are designated by $\phi_{8.23}$, $\dot{\phi}_{4.0}$ and $\ddot{\tau}_{7.0}$, respectively. This particular plot is made for the first bending mode frequency at a time of flight corresponding to maximum dynamic pressure. The basic rate gyro is located aft of the first bending mode antinode. The attitude gyro is located forward of the antinode in the Instrument Unit. The adaptive signal for the first mode was the difference between the output of the basic rate gyro and a rate gyro located in the Instrument Unit. The adaptive signal vector is approximately 180 degrees out of phase with the basic rate signal vector. The vector representing the total open loop gain is the sum of the attitude vector, basic attitude rate vector, accelerometer vector, and the adaptive attitude rate vector. The adaptive gain will tend to make the total vector A lie near the zero degree reference (Fig. 8). The resultant total vector will be approximately the peak value of an open loop gain-phase plot. One can use the vector plot to ascertain how the corresponding gain-phase plot will change when the bending mode parameters vary. Although the phase-sensing system remained satisfactorily stable for slope changes by a factor of four, the system stability margins proved to be sensitive to ±20 percent variations in bending frequency.

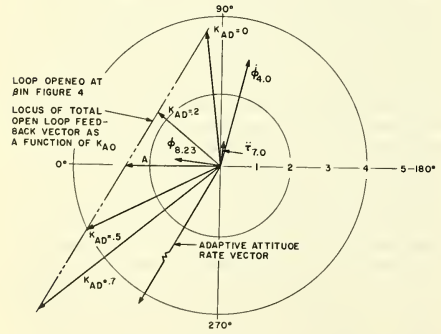


FIGURE 8. VECTOR PLOT OF THE PHASE-SENSING SYSTEM.

Basically, both the modified model reference system and the phase-sensing system are gain adaptive systems. Because stability margins were affected more by variations in bending frequency than by variations in bending mode slopes, it was decided to pursue a phase stabilization system that would adapt to frequency variations.

Adaptive Tracking Offset Notch Filter System

The adaptive tracking offset notch (ATON) filter system is shown in Figure 9. The offset filter is a commutated network and is the same as that used in the ATN filter system. The offset filter in series with the gain K_{N1} (where $K_{N1} > 1$) is a tracking bandpass filter with a constant gain factor. This system differs from the phase-sensing and modified model reference systems which employed fixed bandpass filters with variable gain factors. The ATON filter system was preferred over the other systems because it adapts to bending frequency variations, thus eliminating the large phase shift variations experienced with the fixed bandpass filters, F_1 (Figs. 4 and 5). In the ATON filter system, the phase of the total feedback signal at the first bending mode frequency will always be approximately zero degrees. The amount it will deviate from zero degrees depends primarily on the phase versus frequency characteristics around the first bending mode frequency of the engine actuator system and any common filters used for suppressing higher bending mode signals.

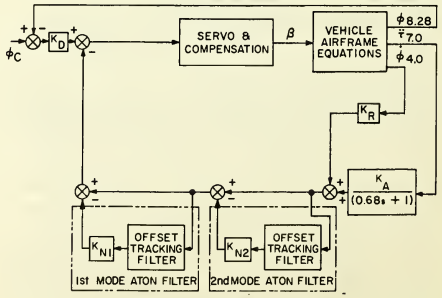


FIGURE 9. BLOCK DIAGRAM OF THE ATON SYSTEM.

This system was not designed to adaptively compensate for mode slope variations. However, by properly selecting the percentage frequency offset and the filter gain K_{N1}, the system can be designed to remain stable for increases in mode slopes of about four and variations in bending frequency of ±20 percent about nominal. The major tradeoff that occurs with this system is between the bending mode damping and the maximum mode slope tolerance. This tradeoff was not as severe as the tradeoffs that occurred in the other three system designs. In using this system, the frequency tracker must be prohibited from tracking modes lower than the first bending mode, i.e., slosh and rigid-body, since this would place additional phase lag in the rigid-body frequency interval. To avoid tracking a lower mode, this system (and also the ATN system) uses a threshold detection circuit which either resets the tracker or makes it continue upward at a predetermined rate whenever the bending mode signal drops below the threshold level. The predetermined rate depends on the bending frequency at liftoff and burnout and the total time of flight. A single-axis miniaturized prototype breadboard of the ATON filter has been built and successfully tested in an analog computer simulation study.

Conclusions

Four adaptive control systems have been studied for stabilizing the first and second bending modes of the Saturn V vehicle. The adaptive tracking notch filter is capable of gain stabilizing the bending modes for the nominal conditions; however, considering off-nominal conditions, the tradeoffs that result with this system make it desirable to seek a different type of design to be used for the first bending mode. The phase-sensing system and the modified model reference system also proved adequate to stabilize the nominal vehicle. These two designs are basically adaptive only to bending slope variations and are not suitable for frequency variations. The adaptive tracking offset notch filter design adequately compensates for the frequency tolerance case where the phase-sensing and modified model reference systems were deficient; however, it has no adaptive adjustment to cope with bending mode slope tolerance variations. This system can, however, be designed to maintain vehicle stability for factors of four increase in the bending mode slope. Based on the results of this study, the adaptive tracking offset notch filter design is preferred over the other three systems for the Saturn V vehicle.

References

1. Hart, W. G., McGowan, G. F., and Vanderberg, F. A.; Advanced Control System for Saturn V Configuration, Phase II Final Report, NASA CR-63-108, November 1963 (Conf.).

2. Hart, W. G., and McGowan, G. F.; Advanced Control System for Saturn V Configuration, Phase III Final Report, NASA CR-64-72, November 1964 (Conf.).

3. Hart, W. G., and McGowan, G. F.; Advanced Control System for Saturn V Configuration, Phase I Final Report, NASA CR-63-24, August 1963 (Conf.).

4. Hosenthien, H. H., and Borelli, M. T.; An Adaptive Tracking Notch Filter for Suppression of Structural Bending Signals of Large Space Vehicles, Astrionics Research and Development Report No. 1, NASA TM X-53000, October 1, 1963.

5. Carroll, S. N.; Analytical Determination of Transfer Functions for RC Commutated Networks, NASA TN D-2948, August 1965.

6. Borelli, M. T., Carroll, S. N., and Hosenthien, H. H.; Recent Developments in Adaptive Tracking Notch Filter Techniques, Astrionics Research and Development Report No. 3, NASA TM X-53217, March 1, 1965.

-- NOTES --

No. 67-592

OPTIMAL CONTROL OF A FLEXIBLE LAUNCH VEHICLE

by

EDMUND G. RYNASKI
Cornell Aeronautical Laboratory
Buffalo, New York

AIAA Paper
No. 67-592

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

7.04, 8.07

OPTIMAL CONTROL OF A FLEXIBLE LAUNCH VEHICLE*

Edmund G. Rynaski
Cornell Aeronautical Laboratory, Inc.
of Cornell University, Buffalo, New York

Abstract

Linear optimal control techniques are used as a synthesis tool to conceptually design a control system for a highly flexible booster. A drift minimum model is used to specify a desirably responding system. The resulting design optimally approximates the response of the model according to a quadratic performance index within the capability of a single control input. Two bending modes are easily damped with a realizable control law. By using matrix transformations, it is shown that two, three or more sensors, of a wide variety of types, located almost anywhere on the vehicle, can be used to obtain the required closed-loop dynamics. The required compensation networks are quickly and easily obtained.

I. Introduction

The post-Saturn class of launch vehicles envisioned by NASA will likely consist of long, slender-bodied boosters. The flexibility of these vehicles cannot be ignored, for the natural frequencies of the normal bending modes will not be significantly higher than the required closed-loop speed of response of the system. Therefore, the flexibility of the booster must be directly considered in the control system design.

This paper describes the application of linear optimal control techniques to the elastic booster problem. These techniques, which guarantee a well-behaved, stable response, show promise as a direct and systematic design procedure that can yield an acceptable control system design. A model is specified that defines a desirable closed-loop response for the system. The minimum of a quadratic performance index is used as a criterion to select a control system design whose response approximates that of the model in a smooth, well-behaved manner and within the available control motion capability.

A Brief Outline of Linear Optimal Theory

The optimal control problem can be briefly stated as follows: starting with some initial condition of the plant, the problem is to find a control motion as a function of time that forces the plant to dynamically respond in a manner that minimizes an integral function of the control and the plant variables. If the plant is linear and the integral contains quadratic functions of the state variables and control variables, the control motion can be generated as a linear combination of the state variables, resulting in a feedback control law. This closed-loop system then responds optimally to any initial condition of the plant. If the plant is describable by a set of constant coefficient linear differential equations of motion and the integral is taken to infinity, the optimal feedback control law is composed of a set of constant feedback gains from the state variables of the plant.

* The work reported upon in this paper was sponsored by the Marshall Space Flight Center of NASA under Contract No. NAS8-20067.

For this problem, the differential equations of booster motion are assumed to have constant coefficients and can be written as a matrix set of first-order equations of the form

$$\dot{x} = Fx + Gu \qquad y = Hx \tag{1}$$

where

- x = the variables of the differential equations of booster motion
- u = the control vector
- y = the output; a transformed set on the state whose motions are to be minimized
- F = an n x n matrix of constants describing the coupling among state variables in the equations of motion
- G = an n x p matrix of constants describing the effect of a control input on the equations of motion
- H = an r x n matrix of constants defining the output of the system.

A model of a desirably responding system is formulated as[3]

$$\dot{y} = Ly$$

where L defines the matrix of constants describing the coupling among variables in the equations of motion of the model.

A control motion u_o is to be found that minimizes the quadratic performance index

$$2V = \int_0^\infty \left[(\dot{y}-Ly)'Q(\dot{y}-Ly) + u'Ru\right] dt \tag{2}$$

where

- Q = an r x r positive definite symmetric matrix whose elements weight the contribution of each error in the performance index
- R = a p x p positive definite symmetric matrix whose elements weight the contribution of each control motion in the integral.

The error portion of the performance index, the $(\dot{y} - Ly)$ term, is a vector that vanishes when the closed-loop optimal system behaves exactly like the model. In general, the feedback cannot exactly match the booster to the model for several reasons. In general, the L and F matrices are of different dimension, and a limited number of controllers is available for feedback purposes.

Although it will be impossible to exactly match the booster and the model, the error can be made as small as practical by weighting the error portion of the performance index heavily with respect to the control. It will be shown that approximating the model response through the use of a quadratic performance index results in not only an optimal, but an acceptable, realizable feedback control system design.

The problem of minimizing Eq. 2 subject to the differential constraint of the equations of motion (Eq. 1) is a straightforward problem in the calculus

of variations. The Lagrangian of the problem can be defined as:

$$\mathcal{L} = \tfrac{1}{2}\left[(\dot{y}-Ly)'Q(\dot{y}-Ly)+u'\mathcal{R}u\right]+\lambda'(-\dot{x}+Fx+Gu)$$

$$= \tfrac{1}{2}\left[x'H'QHx - \dot{x}'H'QLHx - x'H'L'QH\dot{x} + x'L'H'QLHx + u'\mathcal{R}u\right]+\lambda'(-\dot{x}+Fx+Gu) \quad (3)$$

where λ is an n x 1 vector called the adjoint state vector.

The minimization of the quadratic performance index requires that the Euler-Lagrange equations be satisfied.

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = 0$$
$$\frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{u}}\right) = 0 \quad (4)$$

Using the Lagrangian of Eq. 3, the Euler-Lagrange equations are:

$$\dot{\lambda} + F'\lambda - H'QH\ddot{x} + H'QLH\dot{x} - H'L'QH\dot{x} + H'L'QLHx = 0 \quad (a) \quad (5)$$

$$\mathcal{R}u_o + G'\lambda = 0 \quad (b)$$

The optimal control law, from Eq. 5b, is $u_o = -\mathcal{R}^{-1}G'\lambda$ and the optimal closed-loop system becomes

$$\dot{x} - Fx + G\mathcal{R}^{-1}G'\lambda = 0 \quad (6)$$

The exact solution for λ as a function of x has been obtained by Kalman[3] and Tyler[5]. For the case of the single control input, the feedback control law can be more quickly and easily obtained by developing a root square locus expression, which when plotted, spectral-factors the poles of the closed-loop optimal and adjoint systems. The closed-loop characteristic polynomial of the optimal system can then be easily reconstructed from the root square locus plot, and the feedback gains are then easily obtained. The root square locus expression is developed below.

In Laplace transform form, Eq. 5a and 6 become:

$$\begin{bmatrix} [Is-F] & G\mathcal{R}^{-1}G' \\ -H'[-Is-L']Q[Is-L]H & [-Is-F'] \end{bmatrix} \begin{bmatrix} x(s) \\ \lambda(s) \end{bmatrix} = \begin{bmatrix} x(0) \\ -\lambda(0)+H'QH[Isx(0)+x'(0)] \\ -H'[QL+L'Q]Hx(0) \end{bmatrix} \quad (7)$$

The determinant of the left-hand side of Eq. 7, when set to zero, defines the characteristic equation of the optimal system and its adjoint. It can be shown[1] that this determinant may be written

$$\begin{vmatrix} [Is-F] & G\mathcal{R}^{-1}G' \\ -H'[-Is-L']Q[Is-L]H & [-Is-F'] \end{vmatrix} = \mathcal{R}|Is-F||-Is-F'||I+\mathcal{R}^{-1}G'[-Is-F]^{-1}H' \times [-Is-L']Q[Is-L]H[Is-F]^{-1}G| = 0 \quad (8)$$

if $[Is-F]$ and $[-Is-F']$ are square non-singular matrices. But the determinants $|Is-F|$ and $|-Is-F'|$ are the characteristic polynomials of the open-loop system and its adjoint, which vanish only at their root locations. The closed-loop roots of the optimal system and adjoint are therefore obtained from the expression

$$\left|[I+\mathcal{R}^{-1}G'][Is-F']^{-1}H'[-Is-L']Q[Is-L]H[Is-F]^{-1}G\right| = 0 \quad (9)$$

The term $H[Is-F]^{-1}G$ is a matrix of transfer functions $\frac{y_i}{u_j}(s)$ relating the outputs of the system to the inputs and $G'[-Is-F']^{-1}H'$ is the transpose of the transfer function matrix with s replaced by -s. The locus of roots obtained from Eq. 9 defines a root square locus expression[1, 2, 4]. Using this expression, one can investigate the closed-loop characteristic polynomial of the optimal system as a function of the weighting parameters of the performance index (Eq. 2). Because the root square locus is a p x p expression, while the original determinant of Eq. 8 is of dimension 2n x 2n, a considerable simplication has been made for purposes of analysis. In particular, because the booster has a single controller (the gimbaled rocket engines), the root square locus reduces to a scalar expression.

Elastic Booster Application

Because the booster has a single control element, feedback control can alter only the characteristic polynomial of the vehicle and not the zeros of the transfer functions (unless other dynamic elements are deliberately added to the system). Therefore, the optimal control problem may be formulated to express the error in terms of a difference between a differential equation whose characteristic polynomial is that of the elastic booster and a differential equation whose characteristic polynomial is that of the model. To do this, assume that the F matrix associated with the booster is of dimension n x n and L is dimension ℓ x ℓ. Define H

$$H = \begin{bmatrix} h_{11} & 0 & \cdots & 0 \\ 0 & h_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & h_{\ell\ell} \end{bmatrix} \quad (10)$$

where $h_{11} = h_{22} = \ldots = h_{\ell\ell} = 1$.

The matrices F, L, and G are written

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ & & & 1 & 0 \\ 0 & & & 0 & 1 \\ -b_1 & -b_2 & \cdots & -b_{n-1} & -b_n \end{bmatrix}$$

$$L = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & 1 & 0 \\ 0 & & & & & 0 & 1 \\ -d_1 & -d_2 & \cdot & \cdot & \cdot & -d_{\ell-1} & -d_\ell \end{bmatrix}$$

$$G = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} \qquad (11)$$

where

$$s^n + b_n s^{n-1} + b_{n-1} s^{n-2} + \cdots b_2 s + b_1 = |Is - F| \quad \text{(a)}$$

$$s^\ell + d_\ell s^{\ell-1} + d_{\ell-1} s^{\ell-2} + \cdots d_2 s + d_1 = |Is - L| \quad \text{(b)} \qquad (12)$$

The Q matrix is defined as Q = qI, where I is an ℓ x ℓ identity matrix and R = r, a scalar.

When the quantities F, G, H, L, Q and R are substituted into the root square locus expression of Eq. 9, one obtains

$$-1 = \frac{q}{r} \frac{|Is - L||-Is - L'|}{|Is - F||-Is - F'|} \qquad (13)$$

The numerator of Eq. 13 contains the characteristic equation of the model and adjoint while the denominator of Eq. 13 contains the characteristic equation of the booster and its adjoint. The locus of roots of Eq. 13 will therefore originate at the launch vehicle open-loop poles and terminate at the model singularities. The parameter of the locus is q/r, the ratio of the weighting of the error to the control in the performance index.

Equations of Motion of the Elastic Booster

The equations of motion are taken from NASA "Model Vehicle No. 2", which is representative of the post-Saturn vehicles under consideration by NASA. Assuming complete vehicle symmetry, the equations of motion in the x - z plane at t = 78 sec after launch (max q) are given below with certain simplifications. Two normal bending modes are included and the engine actuator is assumed to be described by a first-order differential equation.

pitching equation

$$\ddot{\phi}_e - 0.0733\alpha + 0.45\beta = 0$$

heaving equation

$$-\dot{\phi}_e + 0.0405\phi_e + \dot{\alpha} + 0.01067\alpha + 0.211\beta = 0$$

actuator dynamics (14)

$$\dot{\beta} + 17.9\beta - 17.9\beta_c = 0$$

first normal bending mode equation

$$-5.453\alpha + \ddot{\eta}_1 + 0.02317\dot{\eta}_1 + 5.37\eta_1 - 15.63\beta = 0$$

second normal bending mode equation

$$-2.360\alpha + \ddot{\eta}_2 + 0.05642\dot{\eta}_2 + 31.80\eta_2 - 22.77\beta = 0$$

where
ϕ_e = rigid-body attitude angle
α = angle of attack
β = control deflection angle
η_1 = first normal bending mode variable
η_2 = second normal bending mode variable

The F and G matrices are written in the first-order form of Eq. 11

$$\dot{x} = Fx + Gu \qquad y = Ax$$

and the output matrix A is defined to be an 8 x 8 matrix of constants:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & & & & a_{2n} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{n1} & a_{n2} & & a_{nn-1} & a_{nn} \end{bmatrix} \qquad (15)$$

Then $$\frac{y_i}{u} = \frac{a_{in} s^{n-1} + a_{in-1} s^{n-2} + \cdots a_1}{s^n + b_n s^{n-1} + b_{n-1} s^{n-2} + \cdots b_1}$$

is the transfer function relating the y_i th variable of the equations of motion to the control input. In terms of the numbers of Eq. 14, the first-order equations become:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{x}_7 \\ \dot{x}_8 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -9.074 & 223.5 & -20.71 & -3010 & -193.7 & -665.5 & -38.71 & -17.99 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \beta_c \qquad (16)$$

$$\begin{bmatrix} \dot{\phi}_2 \\ \dot{\phi}_2 \\ \alpha \\ \eta_1 \\ \dot{\eta}_1 \\ \eta_2 \\ \dot{\eta}_2 \\ \beta \end{bmatrix} = \begin{bmatrix} -19.4 & -1376 & -12.6 & -299.4 & -0.755 & -8.055 & 0 & 0 \\ 0 & -19.4 & -1376 & -12.6 & -299.4 & -0.755 & -8.055 & 0 \\ 55.71 & -1375 & -6.062 & -299.8 & -14.33 & -8.085 & -.377 & 0 \\ 83.32 & -2057 & 29.74 & 8946 & 16.95 & 283.4 & 0 & 0 \\ 0 & 83.32 & -2057 & 29.74 & 8946 & 16.95 & 283.4 & 0 \\ 10.63 & -262.5 & 19.42 & 2140 & 12.9 & 407.6 & 0 & 0 \\ 0 & 10.63 & -262.5 & 19.42 & 2140 & 12.9 & 407.6 & 0 \\ 9.074 & -224 & 33.23 & 3008 & 26.66 & 664.1 & 1.616 & 17.9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} \tag{17}$$

Model

The model to be used with this example will be defined by the third-order system

$$\begin{bmatrix} \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \dot{\gamma}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & -1.4 \end{bmatrix} \begin{matrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{matrix} \tag{18}$$

The model incorporates the concept of the drift minimum requirement, the condition that leads to the cancellation of the sum of all force components, such as gust inputs, perpendicular to the nominal flight plane. This cancellation of forces is equivalent to requiring a free integrator in the closed-loop characteristic polynomial. Therefore the model characteristic equation contains a root at the origin of the s plane:

$$|Is-L| = s(s^2 + 1.4s + 1)$$

Analysis of the Optimal System

Substituting the characteristic polynomials of the open-loop launch vehicle and the model along with their adjoints in Eq. 13, the result is

$$0 = 1 - \frac{q}{r} \frac{s^2(s^2 \pm 1.4s + 1)}{(s \pm 17.9)(s \pm .294)(s \pm .242)(s \pm .04175)} \times \left[s^3 \pm 2(.005)2.32s + (2.32)^2\right]\left[s^2 + 2(.005)5.64s + (5.64)^2\right] \tag{19}$$

The locus of roots of the optimal system is plotted in Figure 1. The adjoint, or right-half plane part of the locus is omitted for clarity. The locus shows that the open-loop roots of the booster originally associated with the rigid-body poles tend toward the model singularities, shown in the locus as zeros. The remaining poles, originally associated with the bending modes, tend to become distributed in a Butterworth filter pattern as the parameter of the locus, q/r , becomes large. For any value of the weighting of the error portion of the integral to the control, the accuracy of the approximation to the model in terms of the closed-loop root location can be determined.

Optimal Control Law

The optimal control law is of the form $\beta_c = -Kx$ where K is a 1 x 8 matrix of feedback gains from the state variables to the control inputs. The slode-loop characteristic polynomial is

$$\begin{aligned} \Delta(s) &= |Is - F + GK| \\ &= s^n + (b_n + k_n)s^{n-1} + (b_{n-1} + k_{n-1})s^{n-2} + \cdots (b_2 + k_2)s + (b_1 + k_1) \\ &= s^8 + (17.99 + k_8)s^7 + (38.71 + k_7)s^6 + (665.5 + k_6)s^5 \\ &\quad + (193.7 + k_5)s^4 + (3010 + k_4)s^3 + (20.71 + k_3)s^2 \\ &\quad + (-223.5 + k_2)s + (9.074 + k_1) \end{aligned} \tag{20}$$

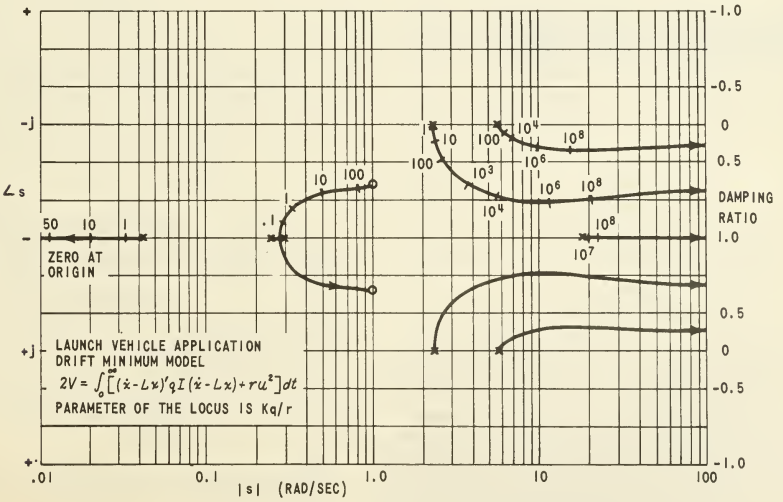


Figure 1 Root Square Locus Plot

67-592

The closed-loop characteristic polynomial can also be obtained directly from the root square locus plot of Figure 1 for any value of q/r. For instance, assuming that $q/r = 100$ will yield an optimal system that results in an acceptable approximation to the model. The closed-loop characteristic polynomial is

$$\begin{aligned}\Delta(s) &= (s+.0077)(s+17.9)\left[s^2+2(.75)(.82)s+(.82)^2\right] \\ &\quad \times\left[s^2+2(.44)(2.6)s+(2.6)^2\right]\left[s^2+2(.12)(6.2)s+(6.2)^2\right] \\ &= s^8+23.02s^7+145.87s^6+1135.8s^5+3372.5s^4+7885.6s^3 \\ &\quad +7157.4s^2+3107.8s+24.11\end{aligned} \tag{21}$$

Equating coefficients of the powers of s in Eq. 20 and 21 yields the feedback control law

$$\begin{aligned}\beta_c &= -15.0x_1-3331.3x_2-7136.7x_3-4875.6x_4-3178.7x_5 \\ &\quad -470.3x_6-107.2x_7-5.03x_8\end{aligned} \tag{22}$$

Or, in terms of the variables of Eq. 17, the control law becomes

$$\begin{aligned}\beta_0 &= -Kx = -KA^{-1}y \\ &= 1.932\phi_R+5.36\dot{\phi}_R+.893\alpha-.239\eta_1-.162\dot{\eta}_1-.456\eta_2 \\ &\quad -.0421\dot{\eta}_2-.281\beta\end{aligned} \tag{23}$$

With the exception of β, the variables of Eq. 23 are still not in terms of directly measurable quantities, i.e., in quantities that directly represent outputs of sensors located on the body of the vehicle. A set of measurable quantities can be chosen, such as the vector

$$z' = \left[\phi_{x_1}, \dot{\phi}_{x_1}, a_{z_{x_1}}, \dot{\phi}_{x_2}, a_{z_{x_2}}, \dot{\phi}_{x_3}, a_{z_{x_3}}, \beta\right] \tag{24}$$

where ϕ_{x_1}, $\dot{\phi}_{x_1}$, and $a_{z_{x_1}}$ represent outputs of a position gyro, a rate gyro, and a normal accelerometer located at station x_1 on the vehicle body.

A similarity transformation $z = By$ can be formulated that defines the measurable quantities of Eq. 24 in terms of the variables of Eq. 23. With this transformation, the feedback control law becomes:

$$\beta_0 = -KA^{-1}y = -KA^{-1}B^{-1}z \tag{25}$$

Choosing the vehicle body stations

$$x_1 = 41.5 \text{ meters}$$

$$x_2 = 86 \text{ meters}$$

$$x_3 = 122.5 \text{ meters}$$

the feedback control law becomes

$$\begin{aligned}\beta_c &= 1.933\phi_{x_1}+5.48\dot{\phi}_{x_1}+.0486a_{z_{x_1}}-.0564\dot{\phi}_{x_2}+.0296a_{z_{x_2}} \\ &\quad -.0592\dot{\phi}_{x_3}+.0244a_{z_{x_3}}-0.35\beta\end{aligned} \tag{26}$$

where the units of β/ϕ and $\beta/\dot{\phi}$ are rad/rad and rad/(rad/sec), and the units of β/a_z are rad/(meter/sec^2). The control law of Eq. 26 involves feedback gains more easily mechanized than those of Eq. 23, but either control law yields complete controllability of the system, resulting in the same closed-loop characteristic polynomial and closed-loop transient response. A selected group of time histories of the vehicle dynamic response to a step command input is shown in Figure 2. The responses verify the smoothness of the response and the predicted damping of the modes. The actuator is capable of responding sufficiently fast to achieve the predicted closed-loop dynamics.

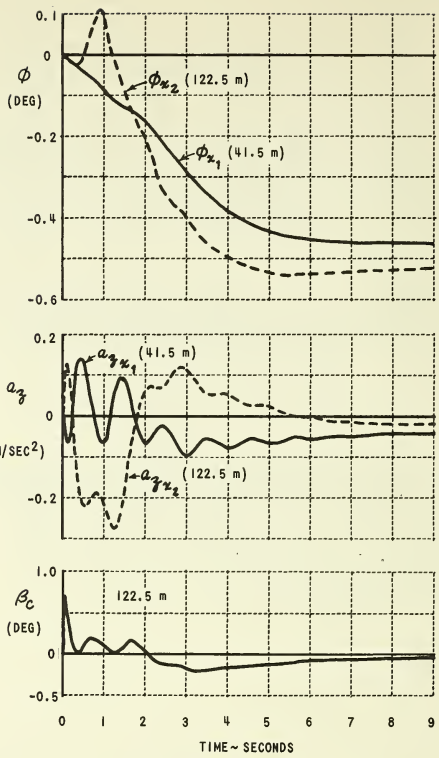


Figure 2 Response of Optimal Regulator

Three-Sensor Optimal Regulator Design

A flow diagram of the desired three-sensor control system is shown in Figure 3.

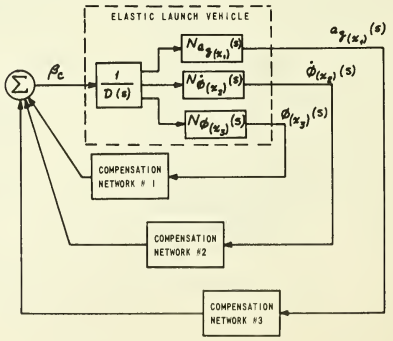


Figure 3 Conceptual Regulator Design

where
$D(s)$ is the open-loop characteristic polynomial, indicated in Eq. 16

$N_{a_{\dot{\gamma}(z_1)}}(s)$ = numerator of the transfer function $\frac{a_{\dot{\gamma}_{z_1}}}{\beta_c}(s)$

$N_{\dot{\phi}_{(z_2)}}(s)$ = numerator of the transfer function $\frac{\dot{\phi}_{z_2}}{\beta_c}(s)$

We wish to describe the control law in the following form:

$$\beta_c = -\left[\ell_1 a_{\dot{\gamma}_{z_1}} + \ell_2 a'_{\dot{\gamma}_{z_1}} + \ell_3 a''_{\dot{\gamma}_{z_1}}\right] - \left[\ell_4 \dot{\phi}_{z_2} + \ell_5 \dot{\phi}'_{z_2}\right] \quad (27)$$
$$-\left[\ell_6 \phi_{z_3} + \ell_7 \phi'_{z_3} + \ell_8 \phi''_{z_3}\right] \quad (27)$$

$$\frac{a'_{\dot{\gamma}_{z_1}}}{a_{\dot{\gamma}_{z_1}}}(s) = \frac{1}{s+\alpha_1} \qquad \frac{a''_{\dot{\gamma}_{z_1}}}{a_{\dot{\gamma}_{z_1}}}(s) = \frac{1}{s+\alpha_2}$$

$$\frac{\dot{\phi}'_{z_2}}{\dot{\phi}_{z_2}}(s) = \frac{1}{s+\beta_1}$$

$$\frac{\phi'_{z_3}}{\phi_{z_3}}(s) = \frac{1}{s+\gamma_1} \qquad \frac{\phi''_{z_3}}{\phi_{z_3}}(s) = \frac{1}{s+\gamma_2}$$

and α_i, β_j and γ_k are defined from

$$N_{a_{\dot{\gamma}_{z_1}}}(s) = a_7 \prod_{i=1}^{7} (s+\alpha_i)$$

$$N_{\dot{\phi}_{z_2}}(s) = b_5 \prod_{j=1}^{5} (s+\beta_j)$$

$$N_{\phi_{z_3}}(s) = c_6 \prod_{k=1}^{6} (s+\gamma_k)$$

The control law of Eq. 27 can then be specified in the form

$$\beta_c(s) = -\left[\ell_1 a_{\dot{\gamma}_{z_1}}(s) + \ell_2 \frac{a'_{\dot{\gamma}_{z_1}}}{a_{\dot{\gamma}_{z_1}}}(s) + \ell_3 \frac{a''_{\dot{\gamma}_{z_1}}}{a_{\dot{\gamma}_{z_1}}}(s)\right]$$
$$-\left[\ell_4 \dot{\phi}_{z_2}(s) + \ell_5 \frac{\dot{\phi}'_{z_2}}{\dot{\phi}_{z_2}}(s)\right] \quad (28)$$
$$-\left[\ell_6 \phi_{z_3}(s) + \ell_7 \frac{\phi'_{z_3}}{\phi_{z_3}}(s) + \ell_8 \frac{\phi''_{z_3}}{\phi_{z_3}}(s)\right]$$

or

$$\beta_c(s) = -\left[\ell_1 + \frac{\ell_2}{s+\alpha_1} + \frac{\ell_3}{s+\alpha_2}\right] a_{\dot{\gamma}_{z_1}}(s) - \left[\ell_4 + \frac{\ell_5}{s+\beta_1}\right] \dot{\phi}_{z_2}(s)$$
$$-\left[\ell_6 + \frac{\ell_7}{s+\gamma_1} + \frac{\ell_8}{s+\gamma_2}\right] \phi_{z_3}(s) \quad (29)$$

The feedback gains ℓ_i are found from the computation

$$L = KA^{-1}$$

where K is a 1 x 8 matrix of gains indicated in Eq. 22. The matrix A is obtained in the following manner:

The first row of A is made up of the coefficients of the numerator polynomial of the $N_{a_{\dot{\gamma}(z_1)}}/\beta_c$ (s) transfer function. The second row is obtained from the coefficients of

$$\frac{N_{a_{\dot{\gamma}_{z_1}}}(s)}{s+\alpha_1} = a_7 \prod_{i=1}^{7} \frac{s+\alpha_i}{s+\alpha_1} = a_7 \prod_{i=2}^{7} (s+\alpha_i)$$

and the other rows are obtained in a similar fashion until eight rows have been obtained.

Substituting the actual numbers of this particular example, the result is

$$K = \begin{bmatrix} -15.0 & -3331.3 & -7136.7 & -4875.6 & -3178.7 & -470.3 & -107.2 & -5.03 \end{bmatrix}$$

$$A = \begin{bmatrix} 382.8 & -9450. & -33.7 & 28620.5 & 164.3 & 2131.5 & 1.794 & -141.1 \\ 80.754 & -1971.9 & 401. & 5838.4 & -1173.9 & 683.78 & -141.1 & 0 \\ 652.25 & -17204.1 & 29241.2 & -1057.84 & 2081.3 & 84.65 & -141.1 & 0 \\ 0 & -16.68 & -1443. & -10.69 & 52.65 & .4431 & 23.49 & 0 \\ 0 & -7.08 & -553.77 & -209.44 & -60.47 & 23.49 & 0 & 0 \\ -10.670 & -1591. & -11.72 & 494.9 & -.5146 & -31.48 & 0 & 0 \\ -2.820 & -468.35 & 134.66 & 106.23 & -31.48 & 0 & 0 & 0 \\ -5.380 & -760.49 & 359.12 & 65.12 & -31.48 & 0 & 0 & 0 \end{bmatrix} \quad (30)$$

The resulting control law is shown in Figure 4.

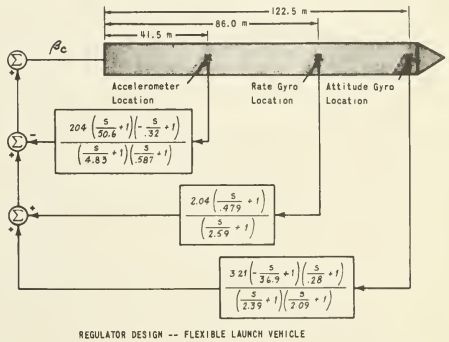


Figure 4 Regulator Design - Flexible Launch Vehicle

A matrix constructed in the fashion defined by Eq. 30 must be square and non-singular. It is not necessarily a similarity transformation representing a definition of quantities that can be sensed on the vehicle. Therefore, there is no guarantee that all the states of the system are completely controllable. We have sacrificed complete control of all the states in order to obtain a simplified control

system configuration. Only the measured quantities, namely, $a_{\dot{z}_{z_1}}$, $\dot{\phi}_{z_2}$, and $\dot{\phi}_{z_3}$ are guaranteed to respond in the manner defined by the optimal, and only then if the initial conditions on the filters are zero. However, since the vehicle is one integral structure, it is felt that complete control of all the states, particularly of the damping of the structural modes have been suitably augmented through feedback. In other words, although the poles of the compensation networks are poles of the system, it can be shown that the response of the measured quantities will contain zero residue in the compensation network poles.

The control system synthesis procedure shown in Figure 4 is not unique. If the launch vehicle characteristic polynomial is of order n, and p measurements of the dynamics are made, then n - p poles for the compensation networks must be chosen. In this particular example, the order of the system is eight and three sensors were used, so five compensation network poles were selected from among the eighteen roots of the numerator polynomials of the transfer functions of $\frac{a_{\dot{z}_{z_1}}}{\beta_c}(s)$, $\frac{\dot{\phi}_{z_2}}{\beta_c}(s)$, and $\frac{\dot{\phi}_{z_3}}{\beta_c}(s)$. Of these eighteen roots, seven right-half plane roots are immediately rejected, for they would give rise to unrealizable compensation networks. Therefore, the search can be narrowed down to the selection of five poles from among eleven candidates. The compensation network poles of Figure 4 represent a deliberate attempt to avoid networks resembling notch filters, simply to show that notch filters do not represent the only solution to the problem.

Conclusions

Linear optimal control techniques, using a model to specify a desirable response, can be used to directly and effectively specify realizable control system designs for large, lfexible launch vehicles. A single root locus plot, used in conjunction with simple matrix techniques, may yield the equivalent of many hours of conventional trial-and-error control system analysis.

References

1. Rynaski, E.G. and R.F. Whitbeck: Theory and Application of Linear Optimal Control. Cornell Aeronautical Laboratory Report No. IH-1943-F-1 (AFFDL-TR-65-28), October 1965.

2. Rynaski, E.G.: The Multivariable Root Square Locus - An Optimal System Design Aid. Presented at the AIAA Guidance and Control Conference, August 1965.

3. Kalman, R.E., T. Englar and R. Bucy: Fundamental Study of Adaptive Control Systems. Vol. I and II, ASD-TR-61-27, March 1961, March 1962.

4. Chang, S.S.L.: Synthesis of Optimum Control Systems. McGraw Hill Book Company, Inc., 1961.

5. Tyler, J.S., Jr.: The Characteristics of Model-Following as Synthesized by Optimal Control. IEEE Transactions on Automatic Control, October 1964.

6. Rynaski, E.G., R.F. Whitbeck, W. W. Wierwille: Optimal Control of a Flexible Launch Vehicle. Cornell Aeronautical Laboratory Report No. IH-2089-F-1, July 1966.

No. 67-593

A MANUAL OPTIMAL GUIDANCE SCHEME USING A PREDICTIVE MODEL

by

J. D. GILCHRIST and D. E. SOLAND
Honeywell, Inc.
Minneapolis, Minnesota

AIAA Paper
No. 67-593

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

2.01, 7.04, 8.08

-- NOTES --

A MANUAL OPTIMAL GUIDANCE SCHEME USING A PREDICTIVE MODEL

James D. Gilchrist, Senior Principal Systems Analyst
Duane E. Soland, Senior Principal Systems Analyst

Systems and Research Division
HONEYWELL INC.
Minneapolis, Minnesota

ABSTRACT

A manual guidance scheme, combining the use of predictive displays and optimal control theory, was investigated for application to launch vehicles during boost. The scheme enables the pilot to continually generate a predicted fuel-optimal trajectory. The problem of determining the optimum steering function is converted into the related two-point boundary value problem. The pilot solves this problem repetitively by adjusting initial conditions on the auxiliary variables so that the predicted vehicle state, as determined by a fast-time model of the vehicle, satisfies the desired end conditions. The resulting trajectory generated by the model is the predicted trajectory for the vehicle. The optimal steering function for the model is converted to real time for steering of the actual vehicle. A particular example was simulated, consisting of boost to a circular orbit. Basic tradeoffs were considered between required accuracy of the fast-time model, accuracy of the resulting end conditions, and pilot work load. Computational and display requirements are discussed, and simulation equations and results are presented.

INTRODUCTION

Guidance and control of vehicles by human operators has always involved some sort of prediction of the controlled variable. In his natural environment, man is capable of performing this task semi-consciously by observing the vehicle and its surrounding environment, as for example, predicting the relative motion of the car he is driving as against that of another vehicle. The complexity and performance of modern aircraft and space vehicles, however, require that a sizeable number of abstract displays be provided for effective use of man in the guidance and control loop.

Up to now, such displays have consisted generally of separate meters registering controlled variable and rate information from which the operator must mentally provide his own prediction for effective control. This is a fatiguing task and relatively inaccurate where guidance of aerospace vehicles is concerned. An alternate approach is the use of predictive displays,[1, 2, 3] where future values of the controlled value are computed electronically and presented to the operator. Generally speaking, such prediction is based on a mathematical model of the system being controlled.

To predict future behavior of a system, it is necessary to assume a time history of the forces which control that behavior. For example, the position and velocity of a spacecraft are determined primarily by the thrust magnitude and direction during the entire launch phase. Also, it is necessary to choose those forces in such a way that some performance criteria is satisfied; e.g., rendezvous with another space vehicle with minimum expenditure of fuel or reentry and subsequent soft landing with minimum total heating. A real-time manual guidance and control problem can be solved by a combination of predictive display[1, 2] and optimal control theory.[4, 5] By this means, the original optimization problem becomes an initial value problem, reducing the manual control requirements to choice of a small number of parameters so that the predicted trajectory intersects the target point in state space. These parameters are the initial values of the auxiliary variables introduced through the optimization theory. In practice, once the correct values of these parameters are found, it is only necessary to introduce small corrections from time to time to account for perturbing effects. Thus, results from optimal control theory are combined with predictive displays to enable man to guide a vehicle along an optimum path to a specified target.

This paper presents the results of applying the predictive model display concept to a particular problem -- that of manual guidance of a launch vehicle from staging to orbit injection. The primary objectives of the study were: (1) to define a manual, optimal, closed-loop guidance scheme which can be used in a variety of missions; (2) to determine the operator's role in the schemes and computer and display requirements; (3) through simulation of the example problem, to determine the accuracy of the desired terminal conditions versus pilot workload; and (4) to compare the simulation results with those of a more conventional manual guidance approach.

* The results reported here were obtained during an investigation carried out under Contract No. NAS 8-20023, sponsored by the Astrionics Laboratory, Marshall Space Flight Center, National Aeronautics and Space Administration.

DESCRIPTION

Figure 1 shows a block diagram of a predictive model guidance scheme which is manual and generates an optimal trajectory with a minimum of computation and display requirements. The guidance scheme consists of three distinct components - the computer, the operator and the display. This man-computer-display system enables the operator to continually generate an optimal trajectory from the vehicle's present state to the desired terminal state. The state of the vehicle is measured at some instant, and the man-computer-display system determines the thrust direction which will steer the vehicle along an optimum path from that point to the specified terminal conditions.

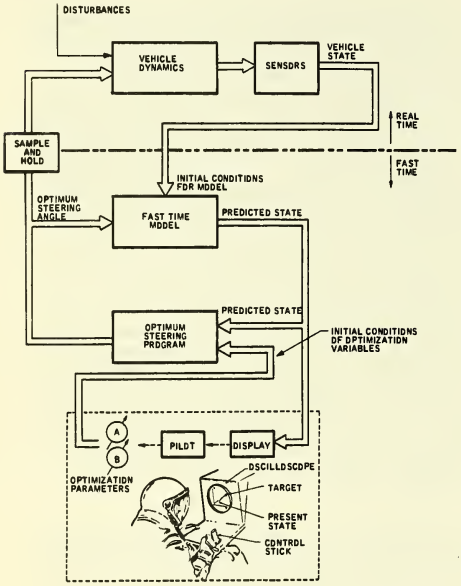


Figure 1. Block Diagram of Predictive Model Guidance Scheme

As the name, "The Predictive Model Guidance Scheme" implies, the basic feature of the scheme is a model of the vehicle which operates in an accelerated time scale. With such a model, the pilot can get an accurate prediction of the future trajectory of the vehicle. The accuracy of the prediction of the vehicle's future response is influenced by the accuracy of the fast-time model as well as by unknown disturbances which may occur at some future time. The fast-time model should be simple in order to minimize the computer speed and computational requirements and yet should generate a fairly accurate predicted trajectory for the real vehicle.

There is a basic tradeoff in the scheme between the operator workload and the computer requirements. As the accuracy of the fast-time model increases, the computer requirements increase, and the operator workload decreases.

The description of the PMGS shown in Figure 1 is divided into two basic parts - the portion which operates in real time t and the portion operating in fast time τ. The block diagram consists of four main sections. The block representing the real-world vehicle dynamics and sensors operates in real time. The complete vehicle present state is assumed to be available from sensor measurements and is the only measurement data required by the fast-time model. A block representing the fast-time model of the vehicle operates in faster than real time and acts as a prediction model of the vehicle dynamics. The initial conditions for the fast-time model are the present state of the real vehicle; an additional optimization program is required for the model. Using results from Pontryagin's Maximum Principle[4, 5] or equivalently the calculus of variations, this optimum steering program consists of an additional set of differential equations describing the auxiliary (adjoint, optimization) variables. The initial conditions of these auxiliary variables (optimization parameters) must be chosen so that the resulting predicted trajectory passes through the specified end points. The operator-display loop permits the proper adjustment of these optimization parameters. On the basis of the fast-time trajectory display, the pilot adjusts the optimization parameters so that the predicted terminal conditions are satisfied. The resulting fast-time steering function is optimum. If it is converted to real time and used to steer the actual vehicle then, in principle, the vehicle would follow an optimal path to the desired terminal conditions. Since the fast-time model operates repetitively, only the initial portion of each optimal steering function is used to steer the vehicle so that no storage of the steering program is required.

Mathematical Description

Figure 2 presents a mathematical description of the PMGS.[4, 5] The vehicle motion is described by

$$\frac{dx}{dt} = F\,[x, u]$$

where x(t), an n-vector represents the vehicle state and u(t), an m-vector, represents the steering function. The ideal model is an exact replica of the vehicle operation in an accelerated time scale,

$$\frac{dx}{d\tau} = F\,[x, u^*],\ x(\tau)_{\tau=o} \text{ and } x(\tau)_{\tau=\tau_f} \text{ given,}$$

where u^* represents an optimal steering function. The auxiliary vector p, an n-vector, is described by

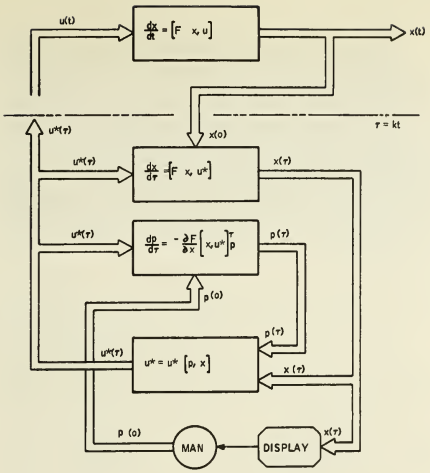


Figure 2. Math. Description of the Predictive Model Guidance Scheme

$$\frac{dp}{d\tau} = \frac{\partial F}{\partial x}[x, u^*]^T p$$

$( )^T$ denotes the transpose. The optimal steering function $u^*(\tau)$ is a function of $x(\tau)$ and $p(\tau)$. The optimal function u^* maximizes $p^T F[x, u]$ where $u \epsilon \Omega$; Ω is the space of allowable steering functions.[2, 3] The resulting two-point boundary value problem is considered solved if p(o) is obtained so that $x(\tau_f)$ is satisfied. The iteration to determine p(o) is performed in the man-display loop. On the basis of the display of $x(\tau)$, man adjusts p(o) until $x(\tau_f)$ is satisfied. Recall that the fast-time model operates repetitively.

Observations

- As the accuracy of the fast-time model increases, the workload of the operator decreases. This implies that an accurate model is desirable. The accuracy of any model, however, is limited by the lack of an exact deterministic model of the vehicle environment and by the basic uncertainty in some of the vehicle parameters. Also, a highly accurate model requires more computation than does a simple approximate model. Thus there exists a tradeoff between operator task loading and computer requirements.

- The operator becomes more effective in the PMGS as the fast-time solution rate increases. For example, the operator is more effective with a rate of one fast-time solution per second than with one solution every 10 seconds. Due to the computation requirements, a finite amount of time is required to compute the fast-time solution. Thus as the solution rate increases, the required computer speed must also increase.

- Much has been said in the literature about the sensitivity of the desired terminal conditions to the optimization parameters. Generally, the sensitivity increases as the required flight time increases. For this reason, the PMGS may not be applicable to all guidance phases, at least not without some modification. This sensitivity consideration is discussed in the following section.

- The number of optimization parameters which the operator must control to guide the vehicle optimally is a definite consideration in the application of the PMGS. For the application to boost guidance, two optimization parameters are required - a number well within the capabilities of an operator.

APPLICATION TO BOOST GUIDANCE

As an example, consider the application of the PMGS to the manual guidance of a launch vehicle during the boost phase. The vehicle is assumed to have two stages, take off horizontally, and develop considerable aerodynamic lift in the first stage. The takeoff weight is approximately 1.5 million pounds, and the initial velocity is nominally 650 ft/sec. The mission profile consists of a planar boost to circular orbit at an altitude of 100 nautical miles, with no restriction on the distance down-range at orbit injection. Variations of lift and drag coefficients with Mach numbers are included in addition to an acceleration limit of 3 g's. The vehicle parameters are characteristic of those of a reusable orbital transport (ROT)[6]. The optimal guidance problem consists of steering the vehicle to the desired orbit with minimum fuel. The minimum fuel problem is equivalent to the minimum time problem, assuming a specified continuous thrust.

From an earlier study,[7] it was determined that five-significant-figure accuracy is required on the optimization parameters to generate an optimum trajectory from first-stage initial conditions to the target conditions. This order of accuracy on the optimization variables yielded errors in the resulting terminal conditions in the order of 10^3 feet in altitude and 10^{-2} degrees in flight-path angle. At the end of the first stage, however, only four-significant-figure accuracy is required to generate an optimal trajectory which yields terminal errors of the same order of magnitude. The required accuracy of the optimization variables decreases as the flight time decreases.

In the hybrid computer simulation used in this study, only three-significant-figure accuracy could be obtained from the analog computer. In view of this sensitivity, the predictive guidance scheme was studied for the second stage only. Even then, the required four-figure accuracy

implied that the least significant figure of the optimization parameters was within the noise level of the analog portion of the hybrid simulation system. The effect of this noise in the analog computer, however, is degrading only in the early portion of the flight; i. e., the sensitivity of the desired terminal conditions to the optimization parameters decreases as the vehicle state approaches the desired terminal conditions.

In the second stage, the flight time is approximately 340 seconds, and the equations of motion are relatively simple. With an optimum choice of coordinate system for the predictive model and appropriate simplifications to the equations of motion, a solution rate of one predictive model solution every second was achieved. The description of the method is divided into the following portions: vehicle dynamics and transformation equations; model dynamics and transformations; optimum steering program and transformations; and finally, the pilot-display link.

Figure 3 shows the equations used in simulating the ROT vehicle dynamics and kinematics during the vacuum phase. Appendix A provides additional information about the model used for the ROT vehicle. Since a different coordinate system is used for the fast-time predictive model, the required transformation equations are also given. These transform the present state of the real-time vehicle in a flight-path coordinate system into initial conditions for the fast-time predictive model in a rectangular coordinate system. In an actual implementation of this guidance scheme, these transformations between coordinate systems would not be required since inertial sensors would be used on board the vehicle.

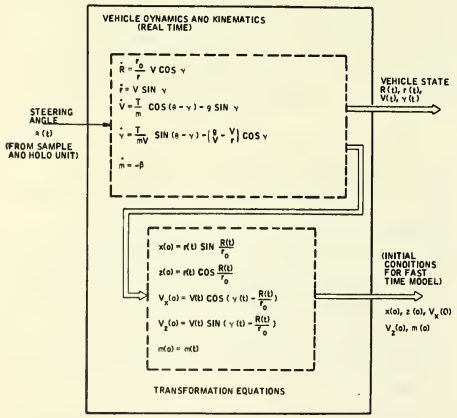


Figure 3. Vehicle Dynamics and Transformations

Figure 4 shows the equations of motion used in the fast-time predictive model. A rectangular coordinate system was used for the predictive model because the equations for the optimization variables uncouple from the equations of motion in such a coordinate system.[5] Since the predicted trajectory is displayed using the same coordinate system as the real-time vehicle model, the required transformation equations are also given. The equations of motion for the fast-time model as well as the optimization equations and transformation equations are presented in Appendix B.

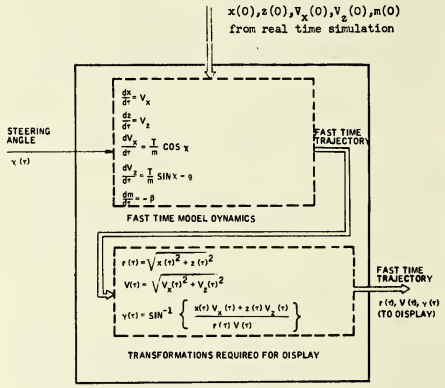


Figure 4. Fast-Time Model Dynamics and Transformations

Figure 5 shows the optimum steering program for the predictive model. In the general case, the equations for the optimization variables (the P_i's) are functions of the state variables; hence, a numerical integration would be required to solve for the optimum steering angle. Due to the rectangular coordinate system chosen for the predictive model, the equations for the optimization

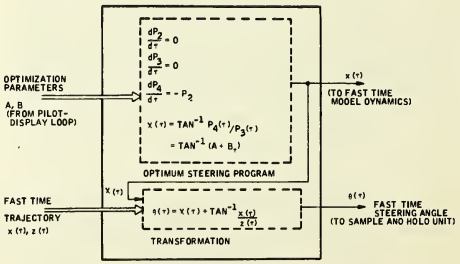


Figure 5. Optimum Steering and Transformations

variables are uncoupled from the state equations and can be solved in closed form. Thus, a time history of the optimal steering function $\chi(\tau)$ is available in closed form. The two parameters A and B defining $\chi(\tau)$ are really the initial conditions for the optimization variables. If a fast-time solution is generated every second, the correct values of A and B required will also change. These changes, however, are predictable. Once correct values for A and B have been found, succeeding values of A are given by the relation

$$A_{new} = A_{old} + B_{old} \text{ (one solution per second assumed)}$$

With this aid, the pilot has only to make minor corrections to A and B due to the inaccuracy of the fast-time model. It is possible to replace the pilot's task of making small corrections to A and B by an automatic system. If only small corrections were necessary, then the increase in computation requirements due to the automatic scheme could be tolerated. However, if large corrections to A and B are required due to unknown disturbances, the required automatic system would be much more complex. As the human pilot is adaptive in nature and can perform the minor corrections or, if required, more major corrections to A and B, thus greatly reducing the computational requirements of a completely automatic system.

Simplifications in the fast-time model were required to yield a closed-form expression for the optimization variables. A flat earth model was assumed, and the constant value for the gravitational acceleration was taken as 31.0. This is the average value of the true value of gravity between the staging and terminal altitude. Further small-angle assumptions were used to reduce the computation time for the predictive model.

Figure 6 shows the operator display loop for the predictive model guidance scheme. As shown, a CRT-type presentation is used for displaying the predicted trajectory. The initial point of the predicted trajectory is the present state of the vehicle. The desired terminal state is displayed on the scope, and the pilot's task is to adjust the parameters A and B so that the predicted trajectory satisfies the desired terminal conditions. To circumvent the scaling problems with such a display, a digital readout presentation (see Figure 6) is also used to give the predicted terminal error in altitude and flight-path angle. The predicted time before velocity cutoff is also displayed. Photographs of the CRT displays used for this guidance scheme are shown in Figures 7 and 8. Notice that the predicted terminal error in Figure 8 is 140 feet on altitude and -0.066 degree on flight-path angle. The symbol o on the figures represents the vehicle present state.

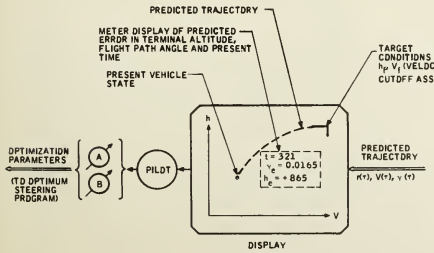


Figure 6. Pilot - Display Link

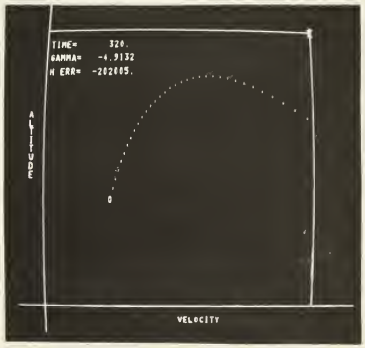


Figure 7. CRT Display - Predicted Trajectory Before Adjustment

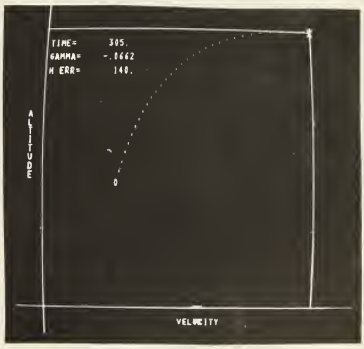


Figure 8. CRT Display - Predicted Trajectory After Adjustment

The operator's task in this guidance scheme is to continually adjust the optimization parameters A and B to minimize the predicted error in the altitude and flight-path angle. In detail, the operations are as follows:

(1) The operator selects values for A and B.

(2) The computer than integrates the predictive model equations of motion and displays the resulting trajectory.

(3) On the basis of the resulting error in the predicted terminal conditions, the operator makes an adjustment to the parameters A and B.

(4) This process is repeated at the rate of one fast-time solution per second until values for A and B are determined which yield zero error in the predicted terminal conditions.

As indicated earlier, two displays are used to aid the operator in his task. The one which gives the operator a display of predicted trajectory in the altitude versus velocity plane is a pursuit display.[8] A pursuit display is defined as containing two moving elements, one representing the actual vehicle state and the other representing the desired state. There is no separate indicator of the error. In this application, however, the pursuit display has only one moving state since the desired terminal state is not a time-varying target. The predicted terminal condition is displayed as a part of the predicted trajectory along with the desired terminal condition. On the basis of the errors between these two terminal conditions, the operator makes an adjustment to the optimization parameters. This type of display is extremely useful in the gross sense, i. e., when the values of A and B are such that the fast-time trajectory is far from the desired. With the display of the predicted trajectory, the operator soon learns how A and B "shape" the trajectory, and hence, how A and B affect the predicted terminal error. A second meter-type display is required to display the predicted terminal errors in the altitude and flight-path angle after the operator has made gross adjustments to the predicted trajectory based on the meter display of the predicted terminal errors. This compensatory display was used rather than scale changes on the pursuit display because of its simplicity. A compensatory display contains one moving element, representing the error in the vehicle state. In this application, this error is the difference in the predicted terminal state and the desired terminal state.

It is possible that an experienced operator could use simply the above compensatory display of the predicted terminal errors. In this simplified display, no CRT presentation is required. Although there are no CRT requirements, very little reduction in computation requirements is expected due to this simplified display. The reason is that the predictive model equations still must be integrated numerically to determine the predicted terminal state. The change in displays is shown in Figure 9. With such a compensatory display, two meters could be used to present the predicted terminal errors to the operator. Although the display requirements can be simplified by this technique, the simplification is not recommended since the operator does not know "where he is" (i. e., the vehicle state) and does not know "where he is going" (i. e., a display of the predicted trajectory). Also, with such a compensatory display, the operator does not have the capability of "shaping the trajectory" with the optimization parameters which he has with the CRT presentation of the predicted trajectory.

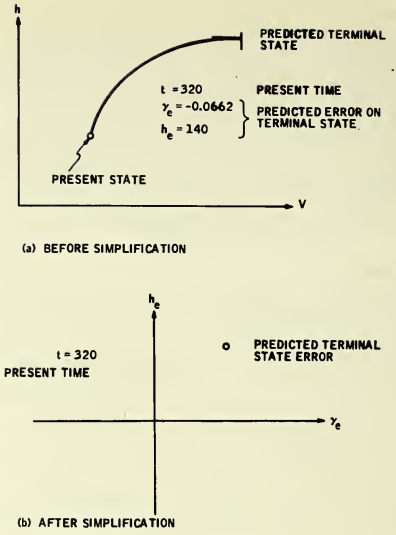


Figure 9. Simplification of the Display for the Predictive Model Guidance Scheme

Results of the Simulation

A series of eight runs with one experienced operator were made to obtain an estimate of the variability of the results which can be expected with the predictive model guidance scheme. Figure 10 shows the trajectories generated and Figure 11 shows the corresponding steering functions. At the beginning of each run the optimization parameters were set to the values which were correct for the nominal second-stage initial conditions. These nominal initial conditions are those which would occur if the first-stage guidance scheme were perfect. The real-time simulation was initiated with off-nominal second-stage initial conditions. The operator was not informed of these initial conditions which differed from the nominal values by approximately ±10 percent.

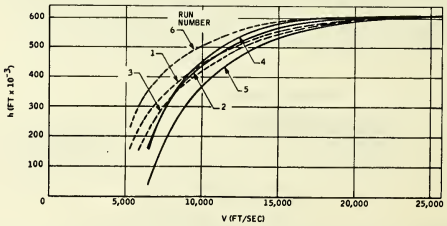


Figure 10. Altitude versus Velocity with Varying Initial Conditions Obtained with the Predictive Model Guidance Scheme

In real time, the operator made major corrections to parameters A and B to account for the off-nominal initial conditions. This period of

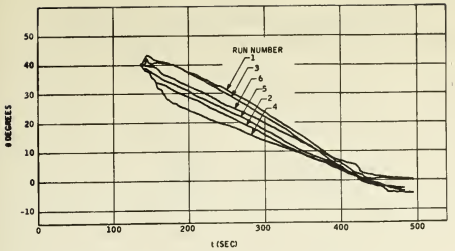


Figure 11. Body Attitude versus Time with Varying Initial Conditions Obtained with the Predictive Model Guidance Scheme

major correction lasted about 30 seconds. At that time, the values of A and B were adjusted so that the predicted trajectory approximately satisfied the terminal conditions. At this time, the approximate predicted errors in the terminal state were ±10,000 feet in altitude and ±0.5 degree in flight path angle. After this period of major corrections until approximately 100 seconds before the predicted cutoff time, the operator monitored the predicted trajectory and, if necessary, made small adjustments to the optimization parameters to keep the predicted errors in the terminal conditions within approximately 10,000 feet in altitude and 0.5 degree in the flight-path angle. At about 100 seconds before cutoff, the operator started making minor corrections to A and B in an effort to null out the predicted terminal errors. At this time, the sensitivity of the predicted terminal conditions to the optimization variables is reduced to the extent that the noise level in the simulation system has no effect on the optimization parameters. Thus, an extremely accurate prediction of the terminal conditions could be utilized. The effect on the steering angle of large corrections at the beginning and small corrections towards the end of the flight are evident in Figure 11. Table 1 presents the results of these runs. To evaluate each trajectory, a measure of terminal error is defined. The performance index, which is a measure of the mean square terminal error, is defined as

$$\text{P.I.} = \sqrt{\frac{\left(\frac{h_e}{h_{e_{max}}}\right)^2 + \left(\frac{\gamma_e}{\gamma_{e_{max}}}\right)^2}{2}}$$

where h_e and γ_e are the actual terminal errors incurred, and $h_{e_{max}}$ and $\gamma_{e_{max}}$ are the maximum errors tolerated in the terminal values of altitude and flight-path angle. In effect, $h_{e_{max}}$ and $\gamma_{e_{max}}$ act as weighting factors for the two errors. No appreciable error for velocity occurred because an automatic velocity cutoff was used in the study. The values used for $h_{e_{max}}$ and $\gamma_{e_{max}}$ were ±20,000 feet and ±0.1 degree. During this experiment, the operator was told to concentrate on the predicted terminal flight-path angle rather than the predicted terminal altitude. The values for $h_{e_{max}}$ and $\gamma_{e_{max}}$ were chosen with this in mind. The rms value of the performance index, altitude error and flight-path angle is 0.075, 1760 feet and 0.007 degree respectively.

COMPARISON WITH MANUAL NOMINAL GUIDANCE SCHEME

To evaluate the PMGS, a second more conventional manual scheme was evaluated.[9, 10] This scheme consisted of a display (altitude versus flight-path angle) of an optimal nominal trajectory. The operator's task was to control the vehicle altitude so as to follow the nominal path. A meter display of the vehicle's present state was also available to the pilot. After sufficient training and many runs, the rms value of the performance index, altitude error and flight-path angle were 0.9, 2200 feet and 0.17 degrees, respectively. An automatic velocity cutoff was used in this simulation.

CONCLUSIONS

The following conclusions are based on simulation results of the PMGS for second-stage boost guidance:

- A human operator is effective in the manual scheme with a fast-time solution rate of one per second.
- A digital readout or meter-type display of the predicted terminal errors in altitude and flight-path angle during the terminal phase of boost is a definite requirement to obtain the desired accuracy in the terminal conditions.
- A two-dimensional display of the predicted trajectory in the altitude-velocity plane is useful to the operator for the iterative task of "shaping" or synthesizing the predicted trajectory. After the trajectory has the proper shape, the meter display is required to yield the desired accuracy in the terminal conditions.
- A simplification of the display requirements toward a meter-type presentation of the predicted terminal errors is not recommended since pilots may desire information about their current status (i.e., the present state of the vehicle) and also information concerning their future flight path (i.e., the predicted trajectory in the altitude-velocity plane).
- Only two optimization parameters are required to manually steer the planar vehicle model to the desired terminal conditions.
- In a mechanization of the predictive scheme, no transformation equations from vehicle to model coordinate system are required since an inertial navigation system could be used to operate in the same reference frame as the fast-time predictive model.

- The operator work load is a function of mission time: initially moderate, then decreasing to zero, and finally, increasing near cutoff conditions.
- The amount of operator training required for efficient operation of the manual guidance scheme is low.
- On the basis of simulation, the accuracy of the PMGS is an order of magnitude better than that obtained with a conventional manual nominal guidance scheme.
- The PMGS appears to be flexible. If fuel permits, a pilot can steer to a new target orbit at any point in the mission, using minimum fuel. Successful simulation runs were made in which the operator was instructed to steer to a 150 mile instead of the nominal 100 mile orbit.
- A CRT-type display for the trajectory and a digital readout of the predicted terminal errors are required.
- Assuming digital computer characteristics similar to a Saturn Launch Vehicle computer,[11] the PMGS computer requirements are: estimated storage capacity - 300 words, estimated computation time - 650 msec every second, three D/A converters to drive the displays, and three A/D converters to input sensor data.

APPENDIX A
REAL-TIME MODEL SIMULATION - PLANAR CASE

A flight-path coordinate system was used in the real-time simulation of the Reuseable Orbital Transport. The force diagram in this coordinate system is shown in Figure A1.

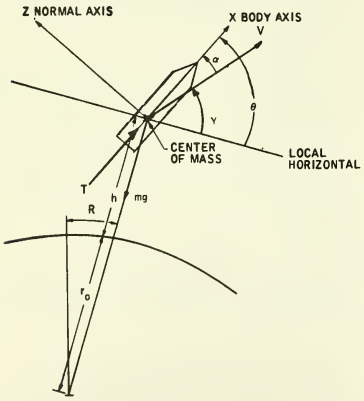


Figure A1. Force Diagram

The equations of motion describing second stage behavior (vacuum phase) are as follows:

$$\dot{R} = \frac{r_o}{r} V \cos\gamma$$

$$\dot{r} = V \sin\gamma$$

$$\dot{V} = \frac{T}{m} \cos(\theta-\gamma) - g \sin\gamma$$

$$\dot{\gamma} = \frac{T}{mV} \sin(\theta-\gamma) - \left(\frac{g}{V} - \frac{V}{r}\right)\cos\gamma$$

$$\dot{m} = -\beta$$

where

$$T = c^* \beta$$

$$\alpha = \theta - \gamma$$

$$h = r - r_o$$

$$g = g_o \left(\frac{r_o^2}{r}\right)$$

The following constants are given

$$c^* = 14,490 \text{ ft/sec}$$

$$\beta = 21.05 \text{ slugs/sec}$$

$$t_s = 136.7 \text{ sec (second-stage ignition)}$$

$$m_2 = 9.1078 \times 10^3 \text{ slugs (initial mass of second-stage)}$$

$$g_o = 32.17 \text{ ft/sec}^2$$

The nominal 100-nm circular target orbit is specified by the terminal values of r, V, and γ

$$r_f = 21.53402 \times 10^6 \text{ ft}$$

$$V_f = 25,570.5 \text{ ft/sec}$$

$$\gamma_f = 0$$

The total load factor is limited to 3 g's. This was satisfied by throttling the engines whenever the constraint was met. This occurs for about 5 seconds at the end of the first stage and for about 30 seconds at the end of the second stage. Otherwise, the maximum value of the mass flow rate B was chosen.

APPENDIX B
FAST TIME MODEL

The fast time model uses a rectangular coordinate system (see Figure B1) and assumes a flat earth.

System Equations

$$\dot{x} = V_x$$

$$\dot{z} = V_z$$

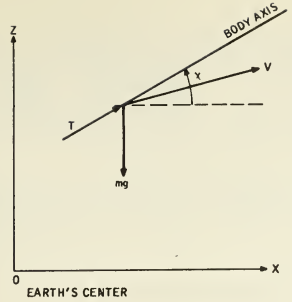


Figure B1. Rectangular Coordinate System

$$\dot{V}_x = \frac{T}{m} \cos \chi$$

$$\dot{V}_z = \frac{T}{m} \sin \chi - g$$

$$\dot{m} = - \beta$$

Auxiliary Equations

The subscripts 1, 2, 3, and 4 refer to x, z, V_x, and V_z respectively.

$$\dot{p}_1 = 0$$

$$\dot{p}_2 = 0$$

$$\dot{p}_3 = -p_1$$

$$\dot{p}_4 = -p_2$$

In writing these adjoint equations, the value of g is assumed constant.

Boundary Conditions

The boundary conditions on the auxiliary variables are:

$$p_1(t_f) = 0$$

$$p_2(t_f),\ p_3(t_f),\ p_4(t_f) \text{ unspecified;}$$

hence, the solutions for the auxiliary variables are:

$$p_1(t) = 0$$

$$p_2(t) = p_{20}$$

$$p_3(t) = p_{30}$$

$$p_4(t) = -p_{20} t + p_{40}$$

where p_{20}, p_{30}, and p_{40} are the unspecified initial conditions.

Optimal Control (minimum time)

$$\chi = \tan^{-1} \frac{p_4(t)}{p_3(t)}$$

$$= \tan^{-1} \left\{ \frac{p_{40}}{p_{30}} - \frac{p_{20} t}{p_{30}} \right\}$$

$$= \tan^{-1} \{ A + B t \}$$

The constants A and B must be chosen so that the resulting solution of the system equations passes through the target conditions.

Target Specifications

$$x^2(t_f) + z^2(t_f) = r_f^2$$

$$V_x^2(t_f) + V_z^2(t_f) = V_f^2$$

$$x(t_f)\ V_x(t_f) + z(t_f)\ V_z(t_f) = 0$$

Transformations

Since the coordinate system for the real-time model differs from this coordinate system for the predictive model, transformations are required to determine the initial conditions for the predictive model. These are:

$$x = r \sin \frac{R}{r_o}$$

$$z = r \cos \frac{R}{r_o}$$

$$V_x = V \cos \left(\gamma - \frac{R}{r_o} \right)$$

$$V_z = V \sin \left(\gamma - \frac{R}{r_o} \right)$$

If the predictive model trajectory is displayed using the coordinate system of the real-time model, the the following additional transformations are required:

$$r = \sqrt{x^2 + z^2}$$

$$V = \sqrt{V_x^2 + V_z^2}$$

$$\gamma = \sin^{-1} \left\{ \frac{x V_x + z V_z}{r V} \right\}$$

$$\theta = \chi + \tan^{-1} \frac{x}{z}$$

Figure B2 shows both the inertial and flight path coordinate systems.

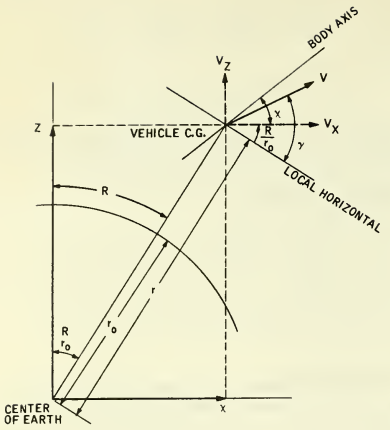


Figure B2. Inertial and Flight Path Coordinate System

BIBLIOGRAPHY

1. Bernstat, R., and Widlok, H., "Principles and Applications of Prediction Display", Journal Inst. of Navigation, Vol. 19, July, 1966, pp. 361-370.

2. Kelley, C. R., "Predictor Instruments Look into the Future", Control Engineering, March 1967, pp. 86-90.

3. Zieboltz, H., and Paynter, F., "Possibilities of a Two-Time-Scale Computing System for Control and Simulation of Dynamic Systems", Proc. National Electronics Conference 9, pp. 215-223 (1954).

4. Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mischenko, E. F., "The Mathematical Theory of Optimal Processes", New York: Interscience Publishers, 1962.

5. Leitmann, G., Editor, "Optimization Techniques", New York: Academic Press, 1962.

6. Lockheed Technical Report, IR 18387, "Advanced Systems - Spacecraft" (December 1964).

7. Bates, J. C., Gilchrist, J. D., Harvester, V. G., and Soland, D. E., "Research on Computational and Display Requirements for Human Control of Space Vehicle Boosters", Honeywell Document 12513-FR1-1 (July 1966).

8. Ely, J. H., Bowen, H. M., and Orlansky, J., "Man-Machine Dynamics", WADC Technical Report 57-582, (November 1957).

9. Hookway, R., et al, "Manual Guidance of Large Space Boosters", Conf. Proc. 1964 National Winter Conv. on Military Electronics, IEE PTGME, February 1964.

10. Williams, D. F., et al, "Manned Booster Control - Man Capabilities", Boeing Document D5-11373, (April 1964).

11. Haeussermann, W., "Guidance and Control of Saturn Launch Vehicles", American Inst. Aero. and Astro. Paper No. 65-304, July 1965.

ACKNOWLEDGEMENTS

The work reported herein was carried out as part of a research project initiated in 1965 and sponsored by the Marshall Space Flight Center under Contract No. NAS 8-20023. This work was directed by J. F. Pavlick of the Marshall Space Flight Center. The authors are grateful to R. C. Kiene, program manager, for his many valuable suggestions and helpful criticisms. The programming efforts of R. Livingston for the hybrid simulation and the contributions of J. C. Bates and V. Harvester on a previous study phase are gratefully acknowledged.

No. 67-594

TERMINAL GUIDANCE FOR SOFT AND ACCURATE MOON LANDING

by

T. Y. FENG and C. A. WASYNCZUK
Hughes Aircraft Company
El Segundo, California

AIAA Paper
No. 67-594

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

-- NOTES --

TERMINAL GUIDANCE FOR SOFT AND ACCURATE MOON LANDING

T. Y. Feng* and C. A. Wasynczuk**
Space Systems Division
Hughes Aircraft Company
El Segundo, California

Abstract

An unmanned lunar spacecraft equipped with terminal descent guidance can achieve a landing accuracy of 500 ft CEP or less. The terminal descent guidance concept proposed in this paper utilizes proportional navigation plus logarithmic deceleration in the initial guided phase of flight and switches at a predetermined range to an unguided gravity turn, thereby achieving a soft landing with vertical attitude at touchdown. Guidance simulation studies have shown this concept to be effective in reducing the target miss distance resulting from dispersions in initial flight conditions caused by uncertainties in retro burnout and errors in instrument measurement and command execution. With the present technology and available spaceflight proven instruments, this guidance concept can be easily implemented without exceeding practical design limitations. Moreover, this guidance concept requires no knowledge of the local vertical throughout the terminal phase of flight. Therefore, the proposed guidance is a feasible concept.

Nomenclature

A = azimuth angle

E = elevation angle

F = thrust vector

F_{ℓ} = proportionality constant

I_{sp} = specific impulse

K = vehicle turning constant

R = range

R^* = range for guidance switching

S_o = navigation constant

V = velocity

W_P = propellant weight

W_i = initial weight

X, Y, Z = inertial coordinates

X_B, Y_B, Z_B = spacecraft body coordinates

a = thrust acceleration

a_c = total command acceleration

a_i = input of thrust acceleration control

a_o = output of thrust acceleration control

$a_{//}$ = acceleration antiparallel to velocity

$a_{\perp}$ = acceleration normal to velocity

g_L = moon gravity constant

h = altitude

$\vec{i}_c$ = unit vector of command thrust acceleration

$\vec{i}_{S/C}$ = unit vector along spacecraft thrust axis

$\vec{i}_v$ = unit vector along spacecraft velocity

$\vec{i}_{\perp}$ = unit vector perpendicular to velocity vector

k = constant in logarithmic deceleration

t = flight time

γ = flight-path angle measured from local horizontal

ϵ = error, angle between desired and actual thrust vector

ξ = velocity vector azimuth angle

η = thrust vector elevation

μ = thrust vector azimuth angle

σ = line-of-sight (LOS) elevation angle

ω = angular velocity

ω_{LOS} = line-of-sight rate

ΔV = thrust increment velocity

subscript o = initial conditions

superscript $*$ = condition at R^*

$\dot{o}$ above symbol = time rate of change

---

*Manager, Systems Analysis.

**Senior Staff Engineer.

67-594

1. Introduction

Future lunar exploratory missions may require logistic support of small payloads on the order of several hundred pounds to be placed on the lunar surface in advance of the astronauts' arrival or to be delivered during the period of lunar surface exploration. For these missions, the payload must be accurately delivered to a preselected location on the lunar surface by an unmanned spacecraft.

Achievement of both a precise and a soft landing of an unmanned spacecraft on the moon's surface will undoubtedly require terminal guidance. Furthermore, if the terminal guidance has to rely on information generated primarily by the onboard instruments, an essentially autonomous guidance system is clearly necessitated.

Various studies of terminal descent have been conducted in the past[1, 2, 3, 4, 5, 6] and have generated a number of guidance schemes and produced partial solutions to the problem. It has been shown[1] that the proportional plus logarithmic navigation type of terminal guidance concept is capable of guiding spacecraft to achieve both a soft and an accurate landing. However, this guidance scheme has no provisions for ascertaining spacecraft vertical attitude at touchdown. Non-vertical landing creates a serious problem of landing stability. On the other hand, a gravity-turn open-loop descent guidance approach[2] employing thrust acceleration in the direction antiparallel to the spacecraft velocity vector provides soft landing with vertical touchdown. Feasibility of this concept was successfully demonstrated by the Surveyor I spacecraft on 2 June 1966.[6] However, the Surveyor gravity-turn descent has no capability to guide the spacecraft to a preselected landing site. A practical solution to terminal guidance of soft-landers that achieves a soft and accurate moon landing with vertical touchdown can be obtained by combining both guidance concepts in an appropriate manner.

The guidance concept presented in this paper is designed to provide an autonomous control of trajectory and velocity of the spacecraft during its terminal descent and to soft-land the spacecraft within a radius of several hundred feet from its predesignated landing site.

2. Basic Requirements

As previously indicated, an unmanned spacecraft on a direct earth-to-moon approach or on a descent from a lunar parking orbit must be provided with terminal guidance if an accurate landing is desired. A typical terminal descent flight (Figure 1) begins right after completion of the main retro phase, which is designed to remove the bulk of spacecraft kinetic energy. Ideally, the retro burnout condition should allow the spacecraft to descend by means of an unguided gravity turn and land on target without miss. In reality, deviations from the ideal condition at retro burnout and errors in instruments and control during the subsequent flight could be large enough to cause untolerable target miss. Therefore, a guidance system is required to correct the spacecraft flight path.

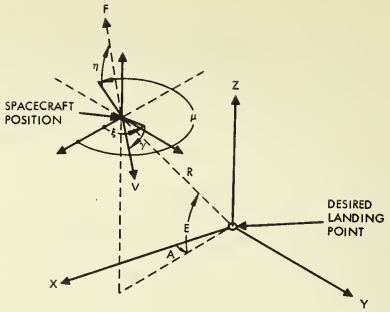


Figure 1. Terminal Approach Geometry

Basic requirements upon such a guidance system are assumed to be as follows:

- Soft landing with impact velocity of not more than 15 fps; this requirement is dictated by spacecraft structural limitations
- Vertical touchdown, to avoid possible spacecraft topple-over at landing
- Landing accuracy of 500 ft CEP at a marked landing point

3. Guidance Concept

The guidance concept presented in this paper combines proportional navigation plus logarithmic deceleration in the initial guided phase of flight with a gravity-turn descent following the initial guided flight to touchdown. The basic principle of this guidance concept is to force the spacecraft in the initial guided phase of flight to descend along a line of sight (LOS) between the spacecraft and target landing point in order to 1) aim at the target and 2) simultaneously decelerate the spacecraft to reduce its kinetic energy. The first objective is achieved by use of the proportional navigation law to generate an acceleration command perpendicular to the spacecraft velocity vector in accordance with $a_{\perp} = S_o (\omega_{LOS} \times V)$, where ω_{LOS} is the LOS angular rate, V is the spacecraft velocity, and S_o is the navigation constant. The second objective is to reduce the spacecraft kinetic energy to achieve a soft landing at the end of the flight. This is accomplished by means of the logarithmic deceleration law, i.e., $dV/dR = K V/R$; based on this law, a thrust deceleration command, antiparallel to the spacecraft velocity vector, is generated in accordance with $a_{//} = F_1 (V^2/R)$, where R is the relative range between spacecraft and the desired landing point, and F_1 is a proportionality factor.

Both S_o and F_1 in acceleration commands are chosen to be constants throughout the flight. For effective control of the spacecraft velocity vector, the value of S_o should be large; while for the noise level and stability considerations, a small value of S_o is preferred. The choice of S_o equal to 3.0 is a design compromise. The magnitude of the proportionality constant, F_1, affects

both the level of the thrust acceleration antiparallel to the spacecraft velocity vector and the terminal descent flight time. The latter relates to the spacecraft fuel consumption which takes the gravity loss ($g_L \times$ flight time) into account. For small values of F_1, say, less than unity, the spacecraft undergoes a small deceleration and descends rapidly toward the landing site. This will shorten the flight time thereby reducing gravity loss. However, the remaining spacecraft kinetic energy after switchover to the gravity turn could be high and thus require an impulsive type deceleration for the spacecraft to attain a soft landing. This may pose a rocket motor design problem. Conversely, if F_1 exceeds 1.5, the spacecraft is subject to a relatively high deceleration in the early phase of guided flight and, consequently, may never reach the guidance switching point of R^*. After several trial and error runs of the computer program, F_1 was selected to be 1.2, based on considerations of fuel economy and spacecraft thrust acceleration capability.

With the value of S_o and F_1 properly chosen, both thrust acceleration terms can be combined vectorially, as follows:

$$\vec{a} = F_1 \frac{V^2}{R} \vec{i}_V + S_o(\omega_{LOS} \times V) \vec{i}_\perp \qquad (1)$$

This equation is used throughout the guided phase of flight until the spacecraft range reaches the value of R^*. R^* is a predetermined relative range between the spacecraft and the target landing site.

At R^*, the flight mode is switched from proportional navigation plus logarithmic deceleration guidance to the unguided gravity turn. The gravity turn utilizes the lunar gravity vector g_L to continuously rotate the spacecraft velocity at a rate of $\dot{\gamma} = + (g_L/V) \cos \gamma$ until the velocity vector coincides with the gravity vector to attain its vertical attitude before or at touchdown. Meanwhile, thrust deceleration proportional to V^2/R is also continuously applied to the spacecraft in a direction antiparallel to the velocity vector to reduce the kinetic energy of the spacecraft to zero at landing. With the gravity turn, the spacecraft achieves a soft landing with vertical attitude at touchdown.

The equations of motion for gravity turn are:

$$\dot{V} = - a + g_L \sin \gamma \qquad (2)$$

$$V \dot{\gamma} = + g_L \cos \gamma \qquad (3)$$

Integrating the above equations and slightly rearranging them gives the following relationship between spacecraft velocity and flight-path angle:

$$\frac{V}{V^*} = \left[\frac{\tan\left(\frac{\pi}{4} - \frac{\gamma}{2}\right)}{\tan\left(\frac{\pi}{4} - \frac{\gamma^*}{2}\right)} \right]^{\left(\frac{a}{g_L} - 1\right)} \frac{\sec^2\left(\frac{\pi}{4} - \frac{\gamma}{2}\right)}{\sec^2\left(\frac{\pi}{4} - \frac{\gamma^*}{2}\right)} \qquad (4)$$

where γ is the flight-path angle measured from the local horizontal, and V^* and γ^* are the velocity and flight-path angle at the guidance switching point, respectively.

The choice of R^* is of paramount importance to landing accuracy. The gravity turn is designed to produce near-optimum performance from the standpoint of the fuel economy. It appears desirable, therefore, to start the gravity turn as early as possible. This implies that a large value of R^* should be chosen. However, since the miss distance of the unguided gravity turn cannot exceed the value of R^*, it is obvious that the smallest R^* possible will result in the most accurate landing. But if a vanishing R^* is used, the spacecraft will soft-land on the target with a probable non-vertical attitude. This is not desirable. Results of simulated flights in a computer study indicate that a maximum R^* of roughly twice the desired landing accuracy is acceptable to achieve satisfactory performance. Based on these considerations, for a desired landing accuracy of 500 ft or less, R^* is chosen to be 800 ft.

4. Guidance Implementation

The guidance concept can be implemented by a conceptual design approach, as shown in the functional block diagram of Figure 2. The guidance system may consist of three strapped-down single-axis rate integrating gyros, a body-mounted accelerometer along the spacecraft vertical axis, a two-axis gimbaled active seeker, a radar altimeter, a doppler velocity sensor, and a special-purpose computer. Spacecraft attitude control and thrust acceleration are derived from the spacecraft flight control system. The thrust vector of the propulsion system is fixed along the spacecraft body vertical axis.

At the beginning of guided flight, the on-board active seeker acquires and locks onto a lunar marker placed at a preselected landing site to generate the required LOS rate and relative range information. In the meantime, the radar altimeter and doppler velocity sensor acquire the moon's surface and generate the necessary spacecraft velocity vector and slant range information. The on-board computer accepts this information and mixes it in a logic circuit, computes it, and generates vehicle steering function for attitude maneuver and thrust magnitude for vehicle deceleration. The attitude maneuver is performed by the flight control system through precision torquing of the strapped-down gyros, while spacecraft deceleration is obtained by means of a propulsion unit in the flight control system. The rate gyros provide the information for stability of the spacecraft, as well as isolation of the seeker-generated inertial LOS rate from the spacecraft body rate. As the spacecraft descends toward the landing site, the guidance system continuously corrects its flight path until it reaches a relative range, R^*. At R^*, a switch is triggered in the guidance computer section which changes guidance from modified proportional navigation to gravity turn. At that instant, the tracking system ceases to function. The radar altimeter and doppler velocity sensor continue to function throughout the gravity-turn phase of flight in which the spacecraft motion is governed by Equation (4).

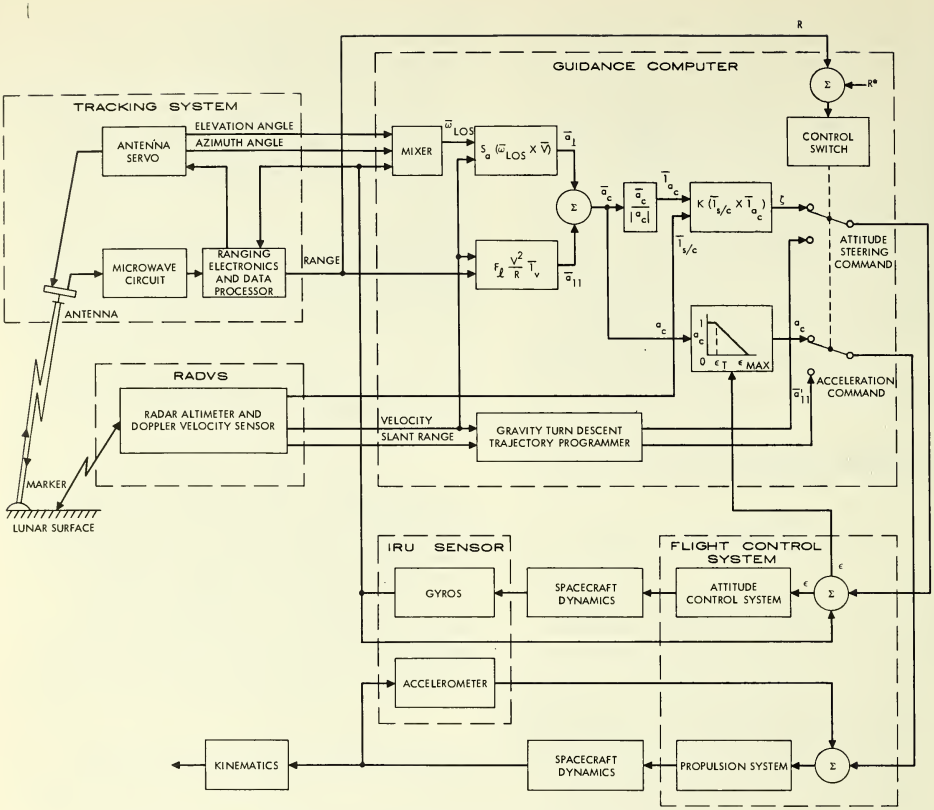


Figure 2. Modified Proportional Navigation Plus Gravity-Turn Guidance System for Accurate Landing of Lunar Soft-Landers

5. Major Assumptions and Control Limitations In Performance Study

A digital computer program has been developed for study of the performance of this guidance concept for lunar terminal descent. Major assumptions in the simulation study are that a flat, non-rotating moon is assumed, based on the large approach flight-path angles, ranging from 60 to 90 deg, with a local horizontal and relatively short flight time of about 200 sec. Because of low initial altitude, ranging from 50,000 through 70,000 ft, a constant lunar gravity of 5.3 ft/sec^2 is also assumed throughout the study.

In mechanizing this guidance concept, the control system responds to guidance commands in two ways. It orients the thrust vector in the desired direction according to the following equation:

$$\vec{i}_c = \frac{\vec{a}_c}{|a_c|} \qquad (5)$$

It also generates the required thrust acceleration level in accordance with Equation (1). However, the total available acceleration of the propulsion system is limited by the capability of the propulsion system. This limit is imposed in such a way that $a_{//}$ is allowed to be executed in full regardless of $a_{\perp}$, i.e.,

$$a_{\perp} \text{ limit} = \sqrt{a_{c\,max}^2 - a_{//}^2} \qquad (6)$$

This limit is imperative for achievement of a soft landing in case the spacecraft is incapable of attaining the desired landing accuracy.

The vehicle turning rate generated by attitude steering command is

$$\vec{\omega}_s = K\left(\vec{i}_{s/c} \times \vec{i}_c\right) \qquad (7)$$

where ω_s is limited to 10 deg/sec maximum based on practical design considerations.

Because of the vehicle turning rate limitation, it may take several seconds to rotate the thrust vector to the proper attitude. Therefore, it is desirable to delay the application of full acceleration command lest the wrong direction thrust should result in an uncontrollable flight trajectory. To avoid that contingency, the thrust acceleration is controlled in the following way:

$$a_o = a_i \qquad \text{for} \quad \epsilon \le \epsilon_T$$
$$a_o = a_i\left(1 - \frac{\epsilon}{\epsilon_{max}}\right) \qquad \text{for} \quad \epsilon > \epsilon_T \tag{8}$$

where ϵ is the error in spacecraft attitude and ϵ_T is the error threshold which is arbitrarily determined.

6. Performance

Performance of a typical soft-lander that employs this guidance concept is examined now in light of landing accuracy, fuel consumption, and guidance performance.

6.1 Landing Accuracy

Landing inaccuracy can be caused by 1) deviation from the nominal initial flight condition at which the guided flight begins and 2) errors in guidance instrumentation and command execution.

Deviation from Nominal Initial Flight Condition. A spacecraft with its nominal initial flight condition will land on target without miss. The nominal initial flight conditions based on the approach geometry shown in Figure 1 are presented in Figure 3 by plotting the initial downrange offset versus flight-path angle for the initial velocity of 500 fps and altitude of 60,000 ft in a plane trajectory case. Actual initial conditions (altitude, downrange to landing target, flight-path angle, and velocity) can depart substantially from these nominal initial conditions. The prime sources of error are uncertainties in the main retro phase performance, which may result in altitude and velocity dispersion, flight-path angle deviation, etc. Altitude dispersion can be caused by uncertainties in the thrust level. Velocity dispersion is due primarily to uncertainties in the total impulse of the retro motor. Flight-path angle deviation could be a result of errors in thrust attitude during the retro phase. When all these uncertainties are accounted for, a positional error of ±10,000 ft (both altitude and downrange) a velocity error of ±100 fps, or a flight-path angle error of ±10 deg, or any combination of these may result.

The effects of these errors on landing accuracy of the spacecraft were studied using a digital computer program. Results of the study are shown in Figures 4 through 9. Based on these results, the target miss appears quite sensitive to errors in downrange, crossrange, and flight-path angle. As shown in Figures 8, 9, and 6, a ±10,000 ft downrange error produced 300 and 270 ft miss, and a typical ±10 deg error in initial flight-path angle would result in 77.5 and 95 ft miss, respectively. The miss resulting from these errors is at least an order of magnitude larger than those produced by errors in altitude, velocity, and azimuth angle.

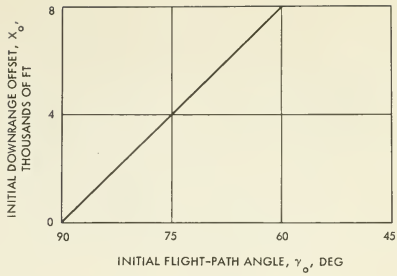


Figure 3. Initial Downrange Offset Distance Versus Flight-Path Angle for Zero Target Miss

Errors in Guidance Instrumentation and Command Execution. Errors in guidance instrumentation are mainly measurement errors in range, velocity, and LOS rate. Range and velocity measurement accuracy are generally functions of range and improve with decreasing range. These errors are not cumulative; because of this, ±2% error in velocity measurement and ±5% error in range measurement caused a miss of only several feet, which is considered insignificant.

An error in LOS rate measurement causes error in thrust acceleration to the spacecraft, resulting in over or undershoot of the target landing site. The source of this error is mainly due to the gyro drift rate. Figure 10 shows a plot of miss versus errors in the LOS rate; a typical 1 mr/sec of LOS rate error would result in a miss of 115 ft.

Execution errors in command pertain primarily to errors of the propulsion system in thrust level and pointing. The thrust acceleration level is controlled by a closed-loop servo. With a high gain control loop, errors could be reduced to a fairly small level; therefore, their effect on landing accuracy is negligibly small. The thrust pointing error caused by the lumped sum errors of the thrust vector misalignment with the spacecraft vertical axis and instrument installation errors can substantially affect miss distance. Figure 11 shows a plot of miss versus thrust pointing error. A typical 0.2 deg pointing error in thrust vector results in a miss of 100 ft maximum.

6.2 Fuel Consumption

Fuel consumption for descent flight to touchdown is expressed in terms of thrust increment velocity, ΔV, given by the following expression:

$$\Delta V = \int_{\Delta t} a_T \, dt \tag{9}$$

where Δt is total flight time from beginning of guided flight to thrust cutoff about 13 ft above touchdown, and a_T is thrust acceleration. The expended thrust velocity increment for three initial flight conditions that result in zero miss is given in Table 1.

The fuel expended during flight in percentage of spacecraft initial weight at the beginning of guided

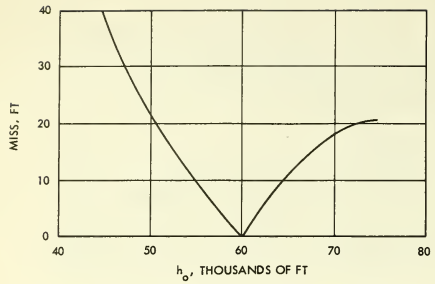


Figure 4. Miss Versus Initial Altitude Deviation (h_o)

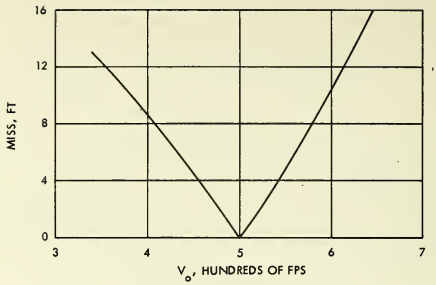


Figure 5. Miss Versus Initial Velocity (V_o)

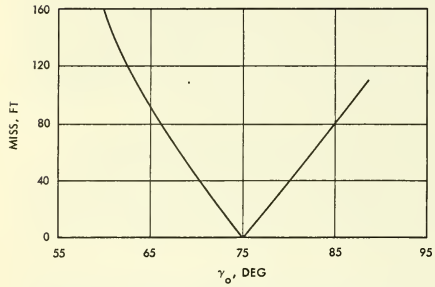


Figure 6. Miss Versus Initial Flight-Path Angle (γ_o)

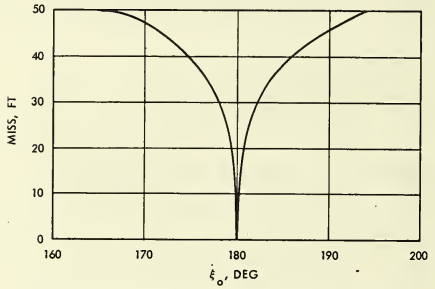


Figure 7. Miss Versus Initial Flight-Path Azimuth Angle (ξ_o)

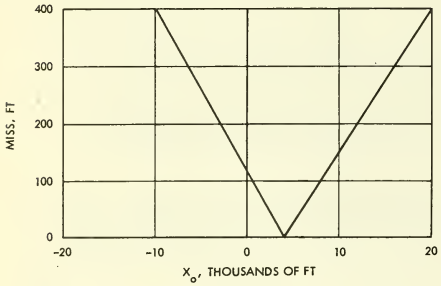


Figure 8. Miss Versus Initial Downrange Offset (X_o)

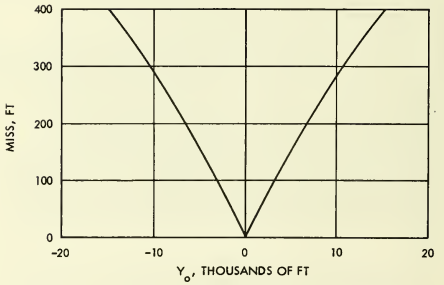


Figure 9. Miss Versus Initial Crossrange Offset (Y_o)

NOMINAL INITIAL FLIGHT CONDITION

V_o = 500 ft/sec γ_o = 75 deg h_o = 60,000 ft X_o = 4000 ft ξ_o = 180 deg Y_o (CROSS RANGE) = 0

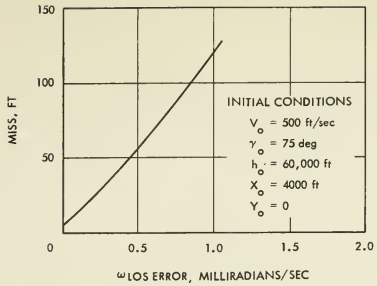


Figure 10. Miss Versus Measurement Error in LOS Rate

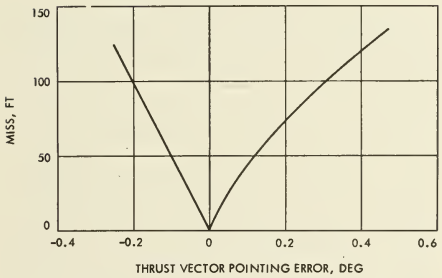


Figure 11. Miss Versus Error in Thrust Vector Pointing

	Units	Condition 1	Condition 2	Condition 3
Velocity, V_o	fps	500	500	500
Altitude, h_o	ft	60,000	60,000	60,000
Flight-path angle, γ_o	deg	90	75	60
Downrange offset, X_o	ft	0	4,000	8,000
Thrust velocity increment, ΔV	fps	1,509	1,515	1,535

TABLE 1. THRUST VELOCITY INCREMENT REQUIRED FOR GUIDED FLIGHT AT THREE NOMINAL INITIAL FLIGHT CONDITIONS

flight $\left(\frac{W_p}{W_i} = 1 - e^{-\Delta V/I_{sp}\,g}\right)$ as a function of the ratio of required thrust velocity increment ΔV to the initial velocity is shown in Figure 12. Even for the 60-deg initial flight-path angle (which demands the largest ΔV), the propellant required is less than 15% of the initial spacecraft weight.

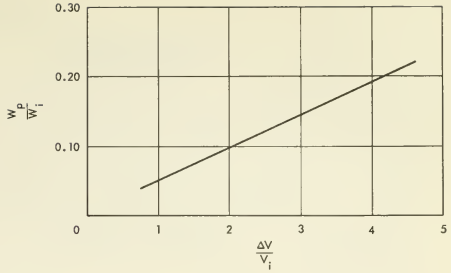
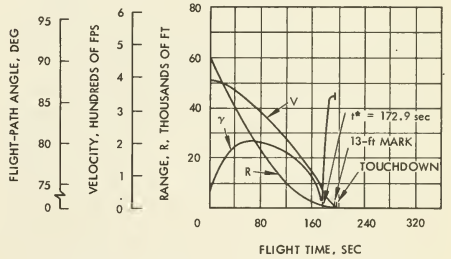


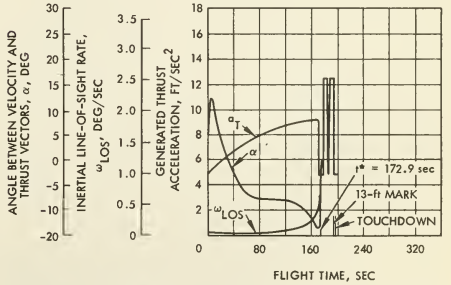
Figure 12. Ratio of Fuel Consumption to Initial Spacecraft Weight Versus Ratio of Thrust Velocity Increment to Initial Spacecraft Velocity

6.3 Guidance Performance

A typical descent flight from an initial condition of 60,000 ft altitude, 500 fps velocity, 75 deg flight-path angle, and a downrange offset of 10,000 ft is given as a sample case for demonstrating system performance. A time history is presented in Figure 13 of the pertinent flight performance parameters: range, velocity, flight-path angle, thrust acceleration, LOS rate, and characteristic angle between velocity and thrust vector. The characteristic angle is the arc tangent of $a_\perp/a_{//}$. The magnitude of this angle depends on the ratio of thrust acceleration normal to velocity to that anti-parallel to velocity.



a) Range, Velocity, and Flight-Path Angle



b) Thrust Acceleration, LOS Rate, and Characteristic Angle Between Velocity and Thrust Vector

Figure 13. Performance Parameters Versus Flight Time

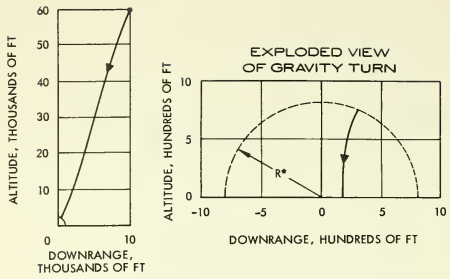


Figure 14. Typical Guided Descent Trajectory

The descent trajectory for this terminal flight is shown in Figure 14. The total flight time is approximately 193 sec to touchdown. The exploded view of the gravity-turn flight within the 800 ft range (Figure 14) clearly indicates that the miss distance of this flight, exclusive of instrumentation and execution errors, is less than 200 ft.

As shown in Figure 13a, velocity and range decrease monotonously with progressing flight time, as expected. The flight-path angle, γ, which starts at 75 deg and ends at 90 deg, experiences considerable variations during the flight. It first increases to about 80.8 deg at the end of 30 sec guided flight. At this instant, the characteristic angle is zero (as shown in Figure 13b), which indicates that resultant thrust acceleration vector is in a direction opposite to the spacecraft velocity vector. Inspection of the trajectory shape in Figure 14 shows that the spacecraft velocity vector at this moment, at the altitude of approximately 45,100 ft, points toward the target landing site; therefore, the LOS rate reaches the minimal (Figure 13b). At a minimum LOS rate, the effect of lunar gravitational force becomes dominant and rotates the spacecraft velocity vector downward. This causes both the LOS rate and the flight-path angle to increase. In the meantime, the characteristic angle, α, changes its sign as it passes the 30-sec flight time mark, as shown in Figure 13b. This phenomenon indicates that the thrust acceleration vector perpendicular to the spacecraft velocity reverses its direction of application. Because a relatively low value of the navigational proportionality constant has been assigned to S_o, the corrective effort to rotate the velocity vector to coincide with the LOS by virtue of thrust acceleration normal to velocity is dominated by the influence of lunar gravity at this portion of the flight. Therefore, the flight-path angle continues to increase until it reaches about 84 deg at approximately 70 sec of flight time. By then, the LOS rate has increased to a level high enough to generate a sufficiently large thrust acceleration normal to the spacecraft velocity to cause the flight-path angle to drop, as shown in Figure 13a. It then continues to decrease to about 73 deg at the time the guidance law switches from proportional navigation to gravity turn at R*. A sharp reverse of trend and rapid increase in γ is shown during the early portion of the gravity-turn descent. This is expected because gravity becomes very influential, while spacecraft velocity is reduced to a relatively low level. Finally, it terminates at 90 deg, as expected, for a vertical touchdown.

Large variations in flight-path angle indicate substantial maneuver requirements of the spacecraft, which, in turn, entail reorientation of the thrust vector and therefore of the spacecraft vertical axis if the body-fixed engines are installed on the spacecraft. The range of variation of γ in this illustrative case does not appear to be excessive and hence is of little concern. However, it should be pointed out that in design installation of the on-board instruments, such as the seeker, radar altimeter, and doppler velocity sensor, the spacecraft attitude maneuver must remain free of the operational limitations of these on-board instruments.

7. Concluding Remarks

The study results have shown that the approach of combining a modified proportional navigation with a gravity turn appeared feasible as a terminal descent guidance for unmanned lunar spacecraft. With this guidance concept and aid of a lunar marker, a spacecraft can achieve a soft and vertical landing on the moon's surface with a landing accuracy of 500 ft CEP or less for moderate deviations in initial flight conditions at the beginning of the guided phase of flight, including the instrumentation and execution errors which are continuously being corrected during flight by the spacecraft guidance system. All of the cases examined indicated that fuel consumption, flight time, and spacecraft attitude maneuver requirements are not excessive when this concept is used. Implementation of this guidance concept is considered relatively easy because most of the on-board instruments and sensors contemplated for the guidance system design are spaceflight proven items.

References

1. Kriegsman, Bernard A. and Reiss, Martin H, "Terminal Guidance and Control Techniques for Soft Lunar Landing," ARS Journal, March 1962.

2. Cheng, Richard K., "Lunar Terminal Guidance," Lunar Missions and Explorations, John Wiley & Sons, Inc., 1964.

3. Citron, Dunin, and Meissinger, "A Terminal Guidance Technique For Lunar Landing," AIAA Journal, March 1964.

4. Green, W.G., "Logarithmic Navigation for Precise Guidance of Space Vehicles," IRE Transactions on Aerospace and Navigational Electronics, June 1961.

5. Hendrix, C.E., "A Soft Landing System Using Optical Sensing Combining Continuously Variable Thrust," Proceedings, National Winter Conference on Military Electronics, February 1962.

6. Surveyor I Mission Report – Part I Mission Description and Performance, Jet Propulsion Laboratory, TR 32-1023, CIT, Pasadena, 31 August 1966.

-- NOTES --

-- NOTES --

No. 67-595

A STEERING LAW SATISFYING THE CONSTANT TOTAL TIME OF FLIGHT CONSTRAINT

by

J. D. CULBERTSON
TRW Systems
Redondo Beach, California

AIAA Paper
No. 67-595

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

8.05, 8.08

A STEERING LAW SATISFYING THE CONSTANT TOTAL TIME OF FLIGHT CONSTRAINT

Jack D. Culbertson
Staff Engineer, Missile Guidance Analysis Section
System Analysis and Software Department
TRW Systems
Redondo Beach, California

ABSTRACT

The problem considered in this paper is that of steering a rocket vehicle so as to achieve a ballistic trajectory which arrives at a specified position at a specified time. A steering equation, which is only a function of the current state of the booster, is derived from the dynamic equation of velocity-to-be-gained, and its accuracy is shown to be relatively insensitive to variations in booster characteristics and to the coordinate system with respect to which steering is done. Moreover, steering can easily be done so as to satisfy such additional constraints and boundary conditions as near optimality with respect to fuel consumption for nonthrottleable, continuous-burn engines, low turning rates, and a roll-axis attitude constraint at cutoff. Simulation results show that, since the steering equation is sufficiently accurate, no trajectory-dependent preflight calculations need be done for it when a simple inflight choice of certain constants is made.

1. INTRODUCTION

This paper both derives and analyzes a method of steering a rocket vehicle so as to achieve a ballistic trajectory which arrives at a specific position at a specific time. Although other methods have been utilized, they often suffer from extensive preflight computations, coordinate system dependence, inaccuracy, and the inability to satisfy other boundary conditions and constraints such as roll-axis attitude at cutoff.

Since the dynamic equation of velocity-to-be-gained is the fundamental equation upon which the steering is predicated, we shall briefly derive it to recall some of its inherent assumptions.

2. THE DYNAMIC EQUATION OF VELOCITY-TO-BE-GAINED

Let us assume that, with respect to some orthogonal, inertial coordinate system, the components of position and velocity of a point mass rocket vehicle are (x_1, x_2, x_3) and $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$, respectively. Also, let a_T be the magnitude of the thrust acceleration, and $\left(g_1(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t),\ g_2(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t),\ g_3(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t)\right)$ be components of the nonthrust acceleration (aerodynamics and gravity) with respect to this coordinate system. The velocity-required is the necessary velocity for an object to hit the target at the constant total time of flight† if a ballistic flight were to begin from a particular position and time. Hence, for a position in the atmosphere, the velocity-required must include the fact that the initial part of the ballistic trajectory will experience atmospheric drag, as may the final part (Figure 1). Let $\left(v_{r1}(x_1, x_2, x_3, t),\ v_{r2}(x_1, x_2, x_3, t),\ v_{r3}(x_1, x_2, x_3, t)\right)$ be the components of the velocity-required for the constant total time of flight constraint. We shall further assume that the components of the velocity-required have continuous first partial derivatives with respect to both position and time.

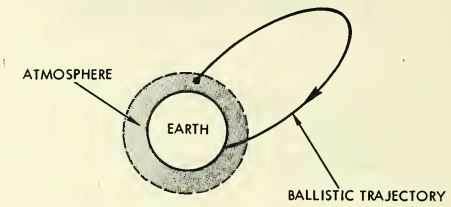


Figure 1. Illustration of Real Ballistics Flight Dynamics

The differential equations of motion in powered flight are

$$\ddot{x}_i = g_i(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t) + a_T \xi_i \qquad i = 1, 2, 3 \qquad (1)$$

where the ξ_i are the direction cosines of the thrust acceleration vector with respect to the reference coordinate axes. The differential equations of ballistic motion are Eq. (1) with $a_T = 0$. The basic problem of steering is to choose the ξ_i during powered flight so that at some position and time

$$\dot{x}_i = v_{ri}(x_1, x_2, x_3, t) \qquad i = 1, 2, 3 \qquad (2)$$

When this occurs, thrust is terminated ($a_T = 0$). For every position and time for which the velocity-required exists (we are, of course, limiting ourselves to such positions and times), the ballistic trajectory defined by that position, time, and the associated velocity-required satisfies Eq. (2).

† This constraint requires that the ballistic trajectory arrive at a specified position (the target) at some time fixed with respect to a reference time such as launch. Clearly, we are considering only targets and total times of flight such that physically realizable ballistic trajectories will satisfy these boundary conditions.

Therefore, using Eq. (1) with $a_T = 0$ and Eq. (2), we obtain after some rearrangement

$$\frac{\partial v_{ri}(x_1, x_2, x_3, t)}{\partial t} = g_i(x_1, x_2, x_3, v_{r1}, v_{r2}, v_{r3}, t)$$

$$- \sum_{j=1}^{3} \left(\frac{\partial v_{ri}(x_1, x_2, x_3, t)}{\partial x_j} \right.$$

$$\left. v_{rj}(x_1, x_2, x_3, t) \right)$$

$$i = 1, 2, 3 \qquad (3)$$

If the velocity-to-be-gained is defined as $v_{gi} = v_{ri} - \dot{x}_i$, $i = 1, 2, 3$, then

$$\dot{v}_{gi} = \dot{v}_{ri} - \ddot{x}_i$$

$$= \sum_{j=1}^{3} \frac{\partial v_{ri}(x_1, x_2, x_3, t)}{\partial x_j} \dot{x}_j + \frac{\partial v_{ri}(x_1, x_2, x_3, t)}{\partial t}$$

$$- g_i(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t) - a_T \xi_i$$

$$i = 1, 2, 3$$

on a powered-flight trajectory. Substituting Eq. (3) for the $\partial v_{ri}/\partial t$ and rearranging yields

$$\dot{v}_{gi} + \sum_{j=1}^{3} \frac{\partial v_{ri}}{\partial x_j} v_{gj} + a_T \xi_i = g_i(x_1, x_2, x_3, v_{r1}, v_{r2}, v_{r3}, t)$$

$$- g_i(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t) \qquad i = 1, 2, 3 \qquad (4)$$

The dynamic equation of velocity-to-be-gained which is usually derived is Eq. (4) with the right-hand side equal to zero.[1] If the rocket vehicle is not in the atmosphere, then the g_i are independent of velocity and the usual equation results as

$$\dot{v}_{gi} + \sum_{j=1}^{3} \frac{\partial v_{ri}}{\partial x_j} v_{gj} + a_T \xi_i = 0 \qquad i = 1, 2, 3 \qquad (5)$$

On occasion we shall refer to the amended version of Eq. (5) by which is meant Eq. (5) with the $\partial v_{ri}/\partial x_j$ assumed to be zero.

It is possible to save the desired form of the dynamic equation even though the vehicle is in the atmosphere by redefining the meaning of velocity-required. This pseudo-velocity-required is defined to be exactly the same as the original except when the position is in the atmosphere on the ascent portion of the trajectory. In this latter case, it is the velocity necessary to hit the target at the constant total time of flight, neglecting the atmospheric drag on the ascent part of the ballistic trajectory. In other words, the ballistic acceleration consists only of gravity on the ascent part of the trajectory (Figure 2). Let the components of this

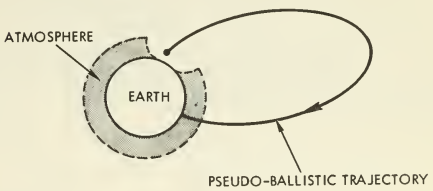


Figure 2. Illustration of Pseudo-Ballistic Flight Dynamics

pseudo-velocity-required be $\left(w_{r1}(x_1, x_2, x_3, t), w_{r2}(x_1, x_2, x_3, t), w_{r3}(x_1, x_2, x_3, t)\right)$. If one now defines $w_{gi} = w_{ri} - \dot{x}_i$ and proceeds to derive a dynamic equation as was done previously, then Eq. (6) results.

$$\dot{w}_{gi} + \sum_{j=1}^{3} \frac{\partial w_{ri}}{\partial x_j} w_{gj} + a_{si} = 0 \qquad i = 1, 2, 3 \qquad (6)$$

where the a_{si} are the components of sensed acceleration (both thrust and aerodynamic). If Eqs. (4) and (6) had been solved in the atmosphere, their solutions would be equal when the trajectory left the atmosphere.

If the reference coordinate system were still orthogonal but earth-fixed, then the right-hand side of Eq. (4) would not vanish out of the atmosphere but would be equal to $2(\underline{\omega} \times \underline{v}_g)_i$ where $\underline{\omega}$ is the earth's angular velocity vector, $\underline{v}_g$ is the velocity-to-be-gained vector, and the subscript i denotes the i^{th} component of the cross product. It would accordingly appear in the yet to be derived steering equation.

If one wished to use the dynamic equation to compute velocity-to-be-gained, then Eq. (6) would be the better method to implement because the $\partial w_{ri}/\partial x_j$ would be easier to compute or approximate and accelerometers measure the a_{si} in any case. However, we shall not use it for this purpose. We are interested in steering outside the atmosphere and shall use Eq. (5) in our analysis and derivation.

One further comment on Eq. (5) is appropriate at this point. Suppose Eq. (1) is accompanied by some steering equation determining the ξ_i. Then the simultaneous solution of these equations will yield the ξ_i and position as functions of time. Under this condition, Eq. (5) is a time-varying, linear differential equation, and, as such, it has

homogeneous and particular solutions which are occasionally referred to in the remainder of this paper.

3. THE STEERING EQUATION

3.1 Derivation

The basic steering problem is now a two-point boundary value problem, i.e., to choose the $\xi_i(t)$ such that the v_{gi} vanish simultaneously (Figure 3). This boundary condition will be called the primary boundary condition. Additional constraints and boundary conditions which may be placed upon steering will be discussed in Subsection 3.3.

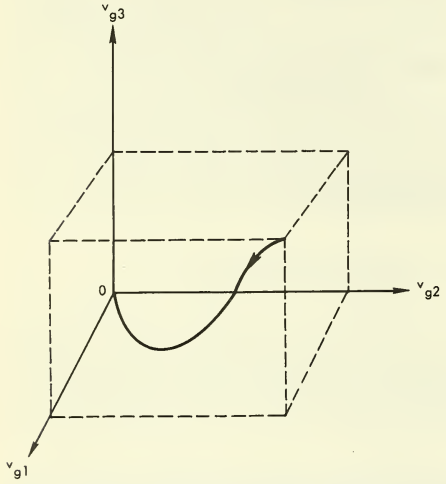


Figure 3. A Trajectory in v_g - Space

There are at least two methods of solving the basic steering problem. The first method is at every guidance cycle to choose $\xi_i(t)$ for the remainder of the powered flight such that the boundary conditions and other constraints are satisfied. This requires an estimate of the thrust acceleration future time history and may be difficult to do. The second method is to choose a solution which satisfies the boundary conditions and other constraints and then try to keep the real solution of Eq. (5) equal to this reference solution. To proceed with the second method, let the f_i, $i = 2, 3$, be homogeneous functions of v_{g1} with continuous first derivatives. These functions do satisfy the primary boundary condition, and, therefore, if we choose the ξ_i so that the real values of v_{gi}, $i = 2, 3$, agree with these functions at every value of v_{g1}, then the velocity-required will be achieved. It is this method which we shall develop below. Of course, the shape of these functions will be determined by practical considerations which will be discussed in Subsections 3.2 and 3.3.

Solving Eq. (5) for ξ_i, $i = 2, 3$, and calling these the commanded values, ξ_{ic}, we obtain

$$\xi_{ic} = -\frac{1}{a_T}\left[\frac{df_i}{dt} + \sum_{j=1}^{3}\frac{\partial v_{ri}}{\partial x_j} v_{gj}\right] \quad i = 2, 3 \qquad (7)$$

assuming $v_{gi} = f_i(v_{g1})$, $i = 2, 3$. If we define $u_i = v_{gi} - f_i(v_{g1})$, $i = 2, 3$, then we want to keep the u_i as small as possible in the presence of missile perturbations. Equation (7) will not do this; thus, it is necessary to add the h_i, functions of the u_i, to Eq. (7) which then becomes

$$\xi_{ic} = -\frac{1}{a_T}\left[\frac{df_i}{dt} - h_i(u_i) + \sum_{j=1}^{3}\frac{\partial v_{ri}}{\partial x_j} v_{gj}\right] \quad i = 2, 3 \qquad (8)$$

Let us now see what some of the characteristics of the h_i must be. Substitution of Eq. (8) into Eq. (5) yields the error equation

$$\dot{u}_i + h_i(u_i) = 0 \qquad i = 2, 3 \qquad (9)$$

Hence, if

$$u_i(t_o) = 0 \qquad i = 2, 3 \qquad (10)$$

where t_o is the time at which this type of steering begins, then Eq. (9) should have the solution $u_i(t) \equiv 0$ because otherwise the steering equation would drive the vehicle off the reference solution even though no missile perturbations were present. We shall require the h_i to have this property.

Let us observe the effect of some missile perturbations upon the proposed steering equation. First, consider a thrust magnitude perturbation, Δa_T, i.e., $a_T = a_{TN} + \Delta a_T$ where a_{TN} is a nominal value. Then substitution of Eq. (8) into Eq. (5) yields Eq. (9) assuming $\xi_i = \xi_{ic}$. Thus, with Eq. (10) satisfied, a reference v_g – trajectory can be steered and the v_{gi} will all vanish simultaneously although a nominal position-velocity trajectory will not be achieved.

Now suppose a thrust misalignment or a center-of-gravity offset† occurs, i.e., $\xi_i = \xi_{ic} + \epsilon_i$; then substituting as before yields

$$\dot{u}_i + h_i(u_i) = -a_T\epsilon_i \qquad i = 2, 3 \qquad (11)$$

†See Section 4 for a precise definition of ϵ_i.

Thus, an additional characteristic of the h_i must be that the u_i are sufficiently small when $v_{gi} = 0$. The error equation is also useful in studying the effect of certain approximations. For instance, suppose the q_{ij}, which may be constants, are used in Eq. (8) to approximate the $\partial v_{ri}/\partial x_j$. Then the substitution of Eq. (8) into Eq. (5) yields an error equation

$$\dot{u}_i + h_i(u_i) = -\sum_{j=1}^{3}\left[\left(\frac{\partial v_{ri}}{\partial x_j} - q_{ij}\right) v_{gj}\right] \qquad i = 2,3 \qquad (12)$$

assuming $\xi_i = \xi_{ic}$. Thus, the effect of certain approximations can be studied provided solutions of the error equation can be obtained. The error equations for approximations to a_T, v_{gi}, and $\dot{v}_{gi}$ can be derived similarly.

Another method can be used to reduce the effect of the driving function $-a_T\epsilon_i$. If the ϵ_i were known, then the insertion of $a_T\epsilon_i$ into Eq. (8) would yield a steering equation which when substituted into Eq. (5) would give Eq. (9) as an error equation assuming $\xi_i = \xi_{ic} + \epsilon_i$. We may not know how to, or wish to, estimate the ϵ_i, but observe that if the expression for $a_T\epsilon_i$ in Eq. (11) were inserted into Eq. (8), then we would get as an error equation for this steering equation

$$\dot{u}_i + h_i(u_i) = -\frac{a_T\epsilon_i}{2} \qquad i = 2,3$$

assuming $\xi_i = \xi_{ic} + \epsilon_i$. Therefore, we see that if the r_i are any real numbers, $r_i \neq -1$, the steering equation

$$\xi_{ic} = -\frac{1}{a_T}\left[-r_i\dot{v}_{gi} + (r_i+1)\left(\frac{df_i}{dt} - h_i(u_i)\right) + \sum_{j=1}^{3}\frac{\partial v_{ri}}{\partial x_j} v_{gj}\right] \qquad i = 2,3 \qquad (13)$$

yields the error equation

$$\dot{u}_i + h_i(u_i) = -\frac{a_T\epsilon_i}{(r_i+1)} \qquad i = 2,3 \qquad (14)$$

and a reduction in the error should be achieved with sufficiently large positive r_i. It also turns out that the r_i have a similar effect upon the driving function for Eq. (12) and do not alter the independence of the steering equation to thrust magnitude perturbations. In practice, the rotational dynamics of the vehicle will limit the values of the r_i as well as the nature of the h_i. Equation (13) is therefore the basic result of this paper. Moreover, if we group the terms in Eq. (13) as in Eq. (15), we see that the r_i play an important role in steering high turning rates because the derivatives of the u_i enter the steering equation.

$$\xi_{ic} = -\frac{1}{a_T}\left[-r_i\dot{u}_i + \frac{df_i}{dt} - (r_i+1)h_i(u_i) + \sum_{j=1}^{3}\frac{\partial v_{ri}}{\partial x_j} v_{gj}\right] \qquad i = 2,3 \qquad (15)$$

An alternate form of Eq. (13) is

$$\xi_{ic} = -\frac{1}{a_T}\left[-r_i\dot{v}_{gi} + (r_i+1)\left(\frac{df_i}{dt} - h_i(u_i)\right) - \dot{v}_{ri} + g_i\right] \qquad i = 2,3 \qquad (16)$$

since from Eqs. (1) and (5) and the definition of v_{gi}

$$\sum_{j=1}^{3}\frac{\partial v_{ri}}{\partial x_j} v_{gj} = -\dot{v}_{ri} + g_i \qquad i = 1,2,3$$

This form is particularly useful in explicit guidance[2] but applicable to delta guidance[2] as well.

Thus far we have not mentioned how ξ_{1c} is to be calculated. Ideally we could solve for ξ_{1c} in Eq. (5), but approximations on the right-hand side, a thrust misalignment, or a center-of-gravity offset might yield a value of ξ_{1c} too inconsistent with the orthonormality constraint. Thus, we take

$$|\xi_{1c}| = \left(1 - \xi_{2c}^2 - \xi_{3c}^2\right)^{1/2}$$

$$\xi_{1c} = |\xi_{1c}|\ \mathrm{sgn}\left[-\dot{v}_{g1} - \sum_{j=1}^{3}\frac{\partial v_{r1}}{\partial x_j} v_{gj}\right]$$

where $\mathrm{sgn}(z) = (z/|z|)$.

3.2 Analysis

An important question is whether the ξ_{ic} will change much during powered flight. This cannot be answered completely, but some simplified analysis

and experience will show that the $\dot{\xi}_{ic}$ are quite small depending on the nature of the f_i and the homogeneous solution to Eq. (5). If we assume that the $u_i(t) \equiv 0$ and neglect the $\partial v_{ri}/\partial x_j$ in Eq. (13), then Eq. (13) reduces to

$$\xi_{ic} = -\frac{1}{a_T}\left[\frac{df_i}{dv_{g1}}\dot{v}_{g1}\right] \qquad i = 2, 3$$

which implies

$$\frac{\xi_i}{\xi_1} = \frac{df_i}{dv_{g1}} \qquad i = 2, 3 \qquad (17)$$

assuming $\xi_{ic} = \xi_i$ and using the amended version of Eq. (5). If $f_i(v_{g1}) = c_{i1} v_{g1}$, where the c_{i1} are constants, then it is easy to see that the ξ_i are all constant using the orthonormality condition and Eq. (17). In fact, it is desirable to have the f_i nearly linear anyway (Subsection 3.3) and the homogeneous solution of Eq. (5) is usually small compared with its particular solution for reasonable thrust levels. Thus, we can conclude that for most situations, Eq. (13) yields nearly constant attitude steering.

Some analysis of the behavior of Eq. (13) has already been done in order to complete its derivation. However, greater insight into its behavior can be obtained by choosing specific functions for the h_i, namely

$$h_i(u_i) = K_i u_i + K_i' \int_{t_o}^{t} u_i(\tau)\, d\tau \qquad i = 2, 3 \qquad (18)$$

where the K_i and K_i' are real constants. With this choice, the error equation becomes

$$\dot{u}_i + K_i u_i + K_i' \int_{o}^{t} u_i(\tau)\, d\tau = E_i(t) \qquad i = 2, 3 \qquad (19)$$

where the $E_i(t)$ are forcing functions of the form mentioned previously and where, for convenience and without loss of generality, t_o is assumed to be zero. The homogeneous solution vanishes because of Eq. (10), and, therefore, this choice of the h_i gives the solutions $u_i(t) \equiv 0$ for Eq. (9). If $K_i^2 < 4K_i'$, the characteristic roots are complex, and letting

$$\alpha_i = \left(\left|K_i^2 - 4K_i'\right|\right)^{1/2}$$

yields the particular solution

$$u_{ip}(t) = d_i \int_o^t \left[E_i(t-\tau)\exp\left(\frac{-K_i\tau}{2}\right)\sin\left(\frac{\alpha_i\tau}{2} - \phi_i\right)\right] d\tau \qquad i = 2, 3 \qquad (20)$$

where $\phi_i = \cot^{-1}(K_i/\alpha_i)$, $d_i = -\sin\phi_i$, and $\exp(\gamma)$ denotes e^{γ}. This particular relation between K_i and K_i' appears to be the best since the $u_i(t)$ are proportional to the integral of the $E_i(t-\tau)$ times a damped sinusoid. In particular, many of the errors inherent in approximations necessary to actually implement Eq. (13) become very small near cutoff and are reasonably large at t_o. Equation (20) says that it is errors near cutoff which are important which, in turn, satisfies one's intuition about steering. With the $K_i' = 0$ in Eq. (19) the particular solution would become proportional to the driving function if the latter were constant. Therefore, this is not a very good choice for Eq. (18).

Another very important aspect of Eq. (13) is its relative independence of coordinate systems. Clearly, the error equations are coordinate-system-dependent only insofar as the $E_i(t)$ are (Eq. (19)). Thus, steering can be done with respect to some coordinate system convenient for other purposes such as navigation. Moreover, concepts such as pitch and yaw planes and the associated ordered rotations need no longer be used since we can use direction cosines, although the mathematical relationships between these concepts can easily be derived if the circumstances so require.

3.3 Constraints and Boundary Conditions

There are various constraints and boundary conditions which may be placed upon steering other than the primary boundary condition. One boundary condition which is necessary is Eq. (10). Another one which can be satisfied is a roll-axis attitude constraint at cutoff (Section 4) by requiring Eq. (17) to be valid at $v_{g1} = 0$ which is cutoff. If $u_i(t) \neq 0$ then, in general, Eq. (17) will not be satisfied at cutoff although it may still be close. It is possible to analyze the error in a manner similar to that employed in the preceding section. Even if a specific roll-axis attitude constraint at cutoff is not desired, Eq. (17) is valid (or nearly so) at cutoff anyway so that the f_i ought to be chosen to yield a reasonable attitude.

It is also desirable to steer so that fuel is not excessively wasted. An analysis of the amended version of Eq. (5) quickly yields that appropriate, constant ξ_i give minimum powered flight time which is equivalent to minimum expended fuel for a nonthrottleable, continuous-burn engine. This is essentially true for Eq. (5) (for reasonable thrust levels) because its homogeneous solution is dominated by its particular solution. More sophisticated analyses employing the calculus of variation on Eq. (5) have also upheld this conclusion. This result coupled with the analysis succeeding Eq. (17) says that the f_i ought to be nearly linear functions.

If a specific attitude at cutoff is not desired, then by the above discussion on optimality, Eq. (17) should be replaced by the constraint

$$\frac{df_i}{dv_{g1}} = \frac{v_{gi}(t_o)}{v_{g1}(t_o)} \qquad i = 2, 3 \qquad (21)$$

at cutoff.

Unless at t_o the f_i are functions such that $\xi_{ic} = \xi_i$ (assuming no thrust misalignment or center-of-gravity offset), the booster may be required to go through a large rotation which is generally undesirable. To eliminate such a maneuver, the f_i will be required to satisfy

$$\dot{u}_i(t_o) = 0 \qquad i = 2, 3 \qquad (22)$$

which yield $\xi = \xi_{ic}$ at t_o by inspection of Eq. (13).

A convenient form for each f_i is

$$f_i(v_{g1}) = \sum_{j=0}^{3} c_{ij} v_{g1}^{\,j} \qquad i = 2, 3 \qquad (23)$$

which has four unknown constants. The c_{io} are to be zero so that the primary boundary condition is satisfied. Equation (17) or (21) determines the c_{i1}. The c_{i2} and c_{i3} are determined to simultaneously satisfy Eqs. (10) and (22). The important point is that this choice should be made inflight at t_o, the time at which this type of steering begins, thus eliminating lengthy preflight calculations of the f_i.

4. SIMULATION RESULTS

This section contains the results of simulating the behavior of a typical booster using Eq. (16) as the steering equation for the nonatmospheric portion of flight, with the h_i as in Eq. (18) and with the f_i as in Eq. (23). It is first necessary to describe briefly the dynamic simulation of the booster and its environment to fully comprehend the results (Figure 4).

Since the attitude control system on most boosters controls the orientation of the pitch, yaw, and roll axes (either through pitch, yaw, and roll angles or through the direction cosines of these axes), and since they command engine angles to cause the booster to rotate to, or to remain at, a desired orientation, the ξ_{ic} generated by the steering equation must be related to the roll axis attitude. The simplest implementation is to have the control system interpret the ξ_{ic} as the desired direction cosines of the roll axis, and this was done in the simulation. Thrust misalignments are nonzero engine angles when the control system actually measures zero engine angles, and center-of-gravity

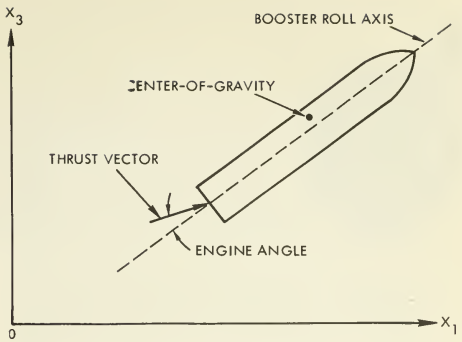


Figure 4. Illustration of Significant Characteristics of Booster Simulation

offsets are the location of the booster center-of-gravity off the roll axis. The rotational dynamics are simulated using the torque with respect to the center-of-gravity generated by the thrust, using its real engine angle, and using appropriate moments of inertia about the center-of-gravity. The commanded engine angle (which may not be the real one because of a misalignment) is related to the roll-axis attitude by an equation of the form

$$\delta_c = c_1(\theta_c - \theta) - c_2\dot{\theta} + c_3 \int_{t_o}^{t} \left[c_1(\theta_c - \theta) - c_2\dot{\theta} \right] d\tau \qquad (24)$$

where δ_c is the commanded engine angle, θ_c and θ are the commanded and measured pitch angles respectively (angle of roll axis in (x_1, x_3)-plane with respect to x_3-axis), $\dot{\theta}$ is the estimated time derivative of θ, and the c_i are non-negative constants. A similar equation is used in the yaw channel. Actuators, or some equivalent system, cause the measured value of the engine angle to be equal to δ_c. If $c_3 = 0$ in Eq. (24) and since the steady-state behavior of the resulting equation (assuming θ_c is constant) is $\dot{\theta} = 0$, then, for a thrust misalignment δ, δ_c must be equal to $-\delta$ which implies $\theta_c \neq \theta$ and thus a roll-axis attitude constraint at cutoff will not, in general, be achieved by Eq. (16). However, with the integral term nonzero it can be shown that the steady-state behavior is $\dot{\theta} = 0$ and $\theta_c = \theta$ for θ_c constant and for a constant thrust misalignment δ since the integral term converges to δ_c. In other words, the integral term remembers the misalignment. A somewhat similar phenomenon occurs for a center-of-gravity offset depending upon the actual nozzle arrangement.

To actually implement the steering equation, it was necessary to make certain approximations. First, the v_{ri} were all approximated by the method of least squares by appropriate polynominals in a suitable region containing the nominal cutoff position

and time. They correspond to a ballistic trajectory which impacts the earth about 5,000 mi away. The $\dot{v}_{ri}$ and the $\dot{v}_{gi}$ were approximated by divided differencing the v_{ri} and the v_{gi} over each guidance cycle, which was 0.25 sec. Here a guidance cycle is the time between issuance of new steering commands to the attitude control system. The value of a_T was approximated by divided differencing the sensed velocity over each guidance cycle. The integration in Eq. (18) was performed trapezoidally. At this point it should be noted that considerable inflight computational simplification occurs in the steering equation if the various items in the above-mentioned approximations are combined. The value of a_T increased monotonically over a steering period of about 45 sec from about 100 ft/sec^2 up to about 300 ft/sec^2. The values of the perturbations in Table 1 were constant over the whole steering period, as were the constants $K_2 = K_3 = 1.0$, $K_2' = K_3' = 0.6$, and $r_2 = r_3 = 0.25$.

As can be seen in Table 1, values of v_{gi}, i = 2, 3, at cutoff are quite small and produce a small miss for all simulated booster characteristics. Note the fact that these values of v_{gi} do not change with booster perturbations, which implies that they occur because of the approximations mentioned above. Although the data are not presented here, it was also found that the guidance cycle length and the approximation to a_T most affected the data in Table 1.

The attitude error is small for the center-of-gravity offset. The importance of the integral in the attitude control system equations was ascertained by not including it for some trial simulations. In these cases, the values of v_{g2}, v_{g3}, and attitude error were significantly affected for thrust misalignments and center-of-gravity offsets. The reason is that the integral eliminates most of the thrust misalignment and much of the center-of-gravity offset prior to cutoff.

5. CONCLUSIONS

A steering equation was derived which is a function of only the current state of the booster and which satisfies the constant total time of flight constraint. It was shown analytically that the accuracy of the steering equation is relatively insensitive to the variation of booster characteristics and to the coordinate system with respect to which steering is done; that the performance can be made near optimum with respect to fuel consumption for non-throttleable, continuous-burn engines; that steering can be done such that small turning rates are encountered; that a roll-axis attitude constraint at cutoff can be achieved; and that the effect of various approximations in the guidance and steering equations can be analyzed. The accuracy of the steering equation as shown in Table 1 and the inflight choice of the constants in Eq. (23) eliminate the need for any trajectory-dependent preflight computations associated with the steering equation.

6. ACKNOWLEDGEMENTS

The results of this paper reflect the efforts of many people, either formerly or currently at TRW Systems, a few of whom can be adequately indicated. Our derivation of Eq. (4) is patterned after J. Bachar's derivation of the dynamic equation of velocity-to-be-gained valid outside the atmosphere, and R. Staley suggested the idea which led to Eq. (6). Equation (8), with the h_i as in Eq. (18), was first arrived at by Bachar by means of a derivation slightly different from ours, although this form of the h_i had been previously employed for other steering equations whose philosophy was somewhat similar to ours. He also recognized the existence of the error equation, Eq. (19), with its particular solution, Eq. (20). The choice of Eq. (23) for the f_i is ancient history although the procedure noted in this paper for determining the constants was suggested by Bachar except for Eqs. (21) and (22), the latter being suggested by D. Phillips. R. Williamson employed some of Staley's results to analyze the optimality of the steering problem by means of the calculus of variations.

REFERENCES

(1) Martin, F. H., "Closed-Loop Near-Optimum Steering for a Class of Space Missions," JACC Preprint, pp 331-344, August 1966.

(2) Pitman, G. R., Jr. (ed), Inertial Guidance, John Wiley and Sons, Inc., New York, pp 244-260, 1962.

Table 1. Simulation Results at Cutoff

	v_{g2} (ft/sec)	v_{g3} (ft/sec)	Attitude Error (deg)	Miss[a] (ft)
Nominal characteristics	-0.02	-0.03	0.00	17
2% thrust magnitude perturbation	-0.02	-0.03	0.00	17
0.2-deg thrust misalignment	-0.02	-0.03	0.00	17
0.1-in. center-of-gravity offset	-0.02	-0.03	0.08	17

[a] The miss is the distance between the actual position on the ballistic trajectory and the target computed at the constant total time of flight and is due entirely to the nonzero values of v_{gi}, i = 2, 3, at cutoff.

No. 67-606

TEST PERFORMANCE OF AN EXPERIMENTAL LASER RADAR FOR RENDEZVOUS AND DOCKING

by

CHARLES L. WYMAN
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-606

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

8.04, 10.06

-- NOTES --

TEST PERFORMANCE OF AN EXPERIMENTAL LASER RADAR FOR RENDEZVOUS AND DOCKING

Charles L. Wyman
Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

A laser radar for rendezvous and docking has been developed over the past several years. The radar uses pulsed injection lasers for acquisition, coarse ranging, and angular tracking. The angular tracking detectors are image dissectors. An incoherent injection diode that is continuously modulated operates as the source for vernier ranging from three kilometers to docking. The radar has recently undergone an extensive test and evaluation program. Static tests in the laboratory indicated a capability of ranging to an accuracy of less than 10 centimeters. Dynamic closed loop testing was performed on a full scale docking simulator, with the radar providing all necessary information to the simulator computer for a complete rendezvous and docking maneuver. The sequence of operation on the simulator was acquisition, angular alignment of the spacecraft, closure, and docking. Several hundred runs on the simulator were made with a successful docking each time. Long range tests were performed from mountain top to aircraft. The tests show that the radar is adaptable to any cooperative rendezvous situation, on manned or unmanned spacecraft, and will provide the necessary intelligence to the guidance computer to effect a complete rendezvous and docking operation automatically.

Introduction

The development of an experimental laser radar for rendezvous and docking began five years ago with inhouse and contractor feasibility studies.[1] The program then progressed to the development of a breadboard radar to demonstrate feasibility and practicality. Once these were established, the development of an experimental prototype was initiated.[2] The prototype approximates a space flight model in size, weight, and power consumption.

The use of lasers and optics for rendezvous and docking provides extensive advantages over present techniques because of the immense increase in antenna gain that is available at optical frequencies. Specifically the diffraction limited angle of any electromagnetic radiator or receiver is

$$\theta = \frac{1.22\lambda}{D} \qquad (1)$$

where λ is the wavelength of the radiation and D is the antenna diameter. Antenna gain derived from this relationship is

$$G = \frac{4\pi A}{\lambda^2} = \frac{\pi^2}{\theta_{hp}^{\ 2}} \qquad (2)$$

where A is the area of the antenna and $\theta_{hp}^{\ 2}$ is the half power angle.[3]

This radar uses gallium arsenide injection lasers, operating at about 9000 Å, and provides many orders of magnitude increase in antenna gain for a given size antenna. The gain may be traded off in a number of ways to decrease size, weight, and power consumption of the radar while maintaining superior performance. The prototype laser radar described here represents the first effort to develop a package in a possible space flight configuration; it has already provided a clear cut indication of size, weight, power, and performance advantages.

System Description and Operation

Performance of rendezvous and docking requires knowledge of the relative positions and velocities of the two spacecraft. The laser radar provides this information to the guidance computer in the form of range, range rate, line of sight angles, and angle rates. In addition, this configuration is capable of acquisition over a 10 degree angle at ranges to 120 kilometers. Table I depicts these parameters and the accuracy of the measurements.

TABLE I. SYSTEM PARAMETERS AND SPECIFICATIONS.

Parameters Range	Long Range 120 to 3 km	Short Range 3 to 0 km
Range accuracy	±0.5%	±0.1 m
Range rate	120 to 50 m/s	50 to 0.3 m/s
Range rate accuracy	±0.2%	±0.03 m/s
Angular position accuracy	±0.1 deg	±0.1 deg
Angular rate accuracy	±0.5 mrad/s	±0.05 mrad/s

The radar is comprised of equipment on each spacecraft. The active vehicle, called the chaser, carries a transceiver that provides range, range rate, angles, and angle rates with respect to the passive vehicle. The passive vehicle, called the target, carries a receiver that measures line of sight angles and angle rates. An acquisition beacon, also on the target vehicle, provides a 10-degree wide beam from an array of gallium arsenide lasers. Table II shows weight and power characteristics of the radar.

67-606

TABLE II. LASER RADAR PHYSICAL CHARACTERISTICS.

System	Power (watts)	Weight (kg)
Chaser		
Optics	22.1	15.15
Electronics		2.90
Display	10.5	2.95
Target		
Optics	12	4.30
Electronics		1.40
Beacon	12	0.52

The prototype laser radar uses both pulsed and cw ranging techniques. Range is generated by transmitting from the chaser and reflecting the beam back from the target by use of optical corner reflectors. Image dissectors are used on both vehicles to generate angular information for acquisition and tracking. Figure 1 shows a block diagram of the system.

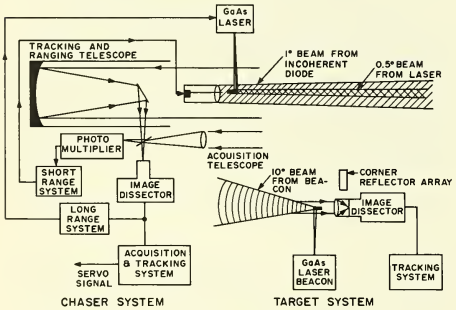


FIGURE 1. BLOCK DIAGRAM OF LASER RADAR.

The acquisition system uses a 10-degree wide laser beacon on the target vehicle and a wide angle telescope and image dissector on the chaser. The image dissector is a photomultiplier in which a small internal aperture is used to limit the active area of the photocathode. By electrostatic focusing and magnetic deflection of the electrons emitted from the photocathode, the active area can be made to scan across the face of the tube. Thus, by means of a quadrature magnetic field, a small instantaneous field of view may be scanned over a larger field of view to generate angular information for acquisition and tracking purposes.

The chaser vehicle contains two transmitters. A pulsed laser transmitter provides a long range capability to an accuracy of 30 meters. In addition, angular tracking is performed with this source. Additional range accuracy is achieved at ranges less than 3 kilometers by a vernier system which uses cw modulation of an incoherent gallium arsenide injection diode and ranges to an accuracy of 10 centimeters. A phototube is used as the detector.

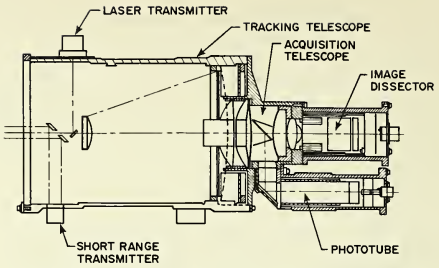


FIGURE 2. CHASER VEHICLE OPTICAL SYSTEM.

The chaser vehicle contains a dual telescope (Fig. 2) in a coaxial configuration. The inner telescope, used for acquisition, has a 10-degree field of view. The outer telescope, used for tracking and ranging, has a 1-degree field of view. The target vehicle has a single telescope that is identical to the inner telescope on the chaser.

The sequence of operation of the laser radar is as follows: Acquisition occurs first; the two spacecraft align to one another with respect to the telescope lines of sight; the target beacon turns off. Tracking on both spacecraft is done with respect to the chaser laser transmitter. Range and angle information is generated and the chaser begins the rendezvous maneuver. When the vehicles are within 3 kilometers, the short range system is turned on. Closure of the two vehicles continues and the vehicles are docked.

For acquisition it is assumed that the two spacecraft are within 120 kilometers of one another and are aligned with respect to the optical axes of the telescope to within ±5 degrees. The 10-degree laser beacon illuminates the chaser and is seen within the 10-degree field of view of the acquisition telescope. This generates angular information causing the chaser vehicle to align itself to the beacon. When this happens, the narrow beam laser from the chaser will illuminate the target. The wide angle telescope on the target vehicle then detects the chaser laser transmitter and will align to it. Simultaneously, the chaser will receive signals from its own transmitter reflected back by the corner reflectors on the target. At this point, tracking on both vehicles may be done with respect to the transmitter on the chaser. The acquisition beacon is no longer needed and is automatically turned off.

The long range system utilizes a double pulse technique that sends out a pair of 100 ns pulses spaced 1 μs apart at a 1 kHz rate. The detection system for the two pulses operates in two modes: a search mode and a track mode. In the search mode, there is no gating and the detector responds to any pulse exceeding a preset threshold. Pulses may occur as a result of the signal, background radiation, or dark current in the detector.

When a pulse is detected, a 67 ns wide gate is generated 1 μs later. If a second pulse occurs during the

gate period, detection is confirmed and the system goes from search mode to track mode. If no confirm pulse occurs, the search mode continues.

Once the track mode takes over, a 500 ns wide tracking gate is generated approximately 1 ms after the confirm gate. When a pulse is detected during the tracking gate period, the confirm gate is again generated 1 μs later.

The occurrence of pulses in a photomultiplier follows a Poisson distribution, and Poisson statistics are applicable to calculation of detection and false alarm probabilities. A false alarm is any sequence of pulses interpreted by the detector as a signal sequence. Clearly as the sequence is lengthened, the probability can be made arbitrarily small by requiring a longer sequence before confirming detection. Figure 3 is a nomogram showing false alarm probability for various ranges,

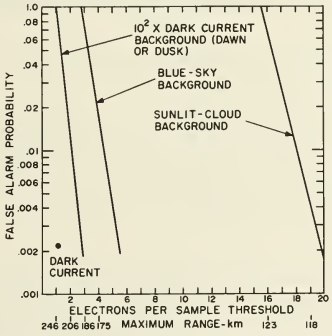


FIGURE 3. FALSE ALARM PROBABILITIES.

backgrounds, and threshold levels and for detection of a single confirmation pulse. The false alarm probability for Figure 3 is the product of the probability of detecting a noise pulse in the search mode and the probability of detecting a noise pulse in a confirm gate period.*

The short range system utilizes a cw modulated injection diode. Conventional phase lock techniques are used. A modulation frequency of 3.75 MHz provides a vernier range accuracy of 0.1 meter with a direct readout of data in meters and tenths of meters.

Test Program

Short Range Tests

Short range tests for calibration of the ranging system were performed in the hallway of the ITT laboratory. Two types of measurements were made over a distance of 22 meters. In one type of measurement, the system was wheeled down the hallway and the distance was measured with a steel tape and compared with the range reading of the laser radar. Readings were recorded every 0.5 meter from 1.5 to 22 meters. In the second type of measurement, data were collected by moving a corner reflector on an optical bench and recording data every centimeter over a distance of 1 meter. These measurements yielded relative range errors over one meter increments.

In the first set of measurements, the range reading was arbitrarily set to read 10 meters at 10 meters and as near as possible to 20 meters at 20 meters. The range was set by changing the length of a cable in the short range system, causing a phase shift with respect to the reference signal.

By arbitrarily setting range, one finds that a mean range error causes a bias in the range readings. Once the bias is known, the range can be reset so that the mean range error is zero. Reduction of the data consisted of establishing the mean range error and the rms deviation from the mean range error. Mean range error and rms deviation from the mean range error were also reduced from the data on the second type of measurements.

Rendezvous Closure and Docking Simulator Tests

The docking simulator used for the tests of the laser rendezvous is a full-scale six-degree-of-freedom system located at the Martin Company in Denver, Colo. The simulator consists of a carriage, which is servo-driven in three translational axes and three rotational axes, in a lightproof room that is 27.4 x 9.7 x 7.3 meters. The servo-drives may be operated on command from a control console or may be controlled from a computer. The computer accepts signals from sensor systems being tested on the carriage and provides a closed loop capability. The carriage represents one vehicle during docking tests, and the second vehicle is fixed on one wall. Figure 4 shows the complete facility, designated as a space operations simulator, with a man mounted on the carriage performing extra vehicular activity tests.

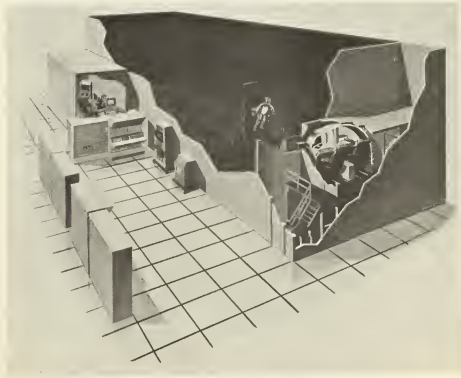


FIGURE 4. SPACE OPERATIONS SIMULATOR.

The laser radar was tested by mounting the chaser vehicle system on the carriage and the target vehicle system on the wall (Fig. 5). The chaser telescope is the white object shown on the carriage in the foreground.

*Information for this discussion was received from Mr. John M. Grant of the International Telegraph and Telephone Corporation.

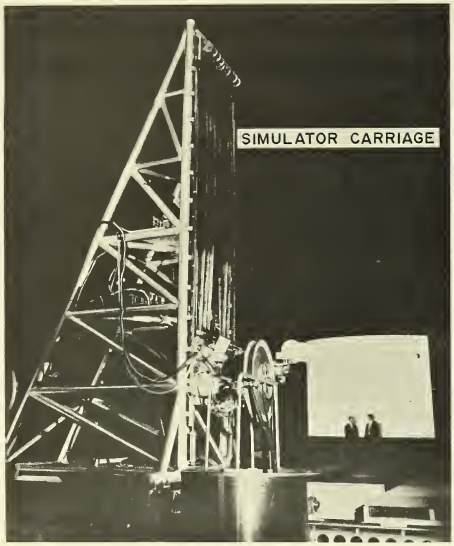


FIGURE 5. OVERALL VIEW OF SIMULATOR TEST SETUP.

The target system is mounted in the upper right-hand corner of the simulated vehicle on the far wall.

A closeup of the mounted chaser vehicle system is shown in Figure 6. The total chaser system is comprised of the telescope, the visual display of range and range rate, and the electronics package.

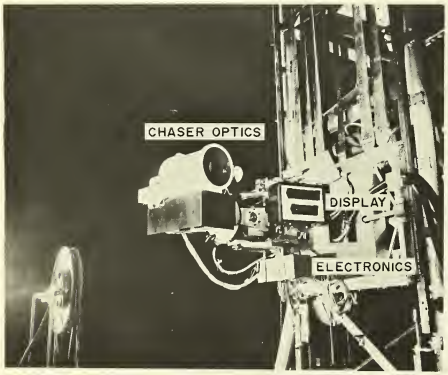


FIGURE 6. CHASER VEHICLE SYSTEM MOUNTED ON THE SIMULATOR.

The target system is shown in Figure 7. The electronic package is out of sight and is about 12.5 centimeters on a side. The ring in the foreground is a docking target located 2 meters from the corner reflector. In the closed

FIGURE 7. TARGET VEHICLE SYSTEM AND TARGET RING.

loop tests, the chaser vehicle system stops in the proximity of the ring and is considered to be docked.

Static testing consisted of commanding the carriage to a certain position, which is compared with the calculated position based on the radar outputs. On-axis tests served to verify the laboratory static tests. Off-axis tests verified beam widths of the various sources and fields of view of each telescope.

Actual closed loop docking was performed during the dynamic tests. Fifteen channels of information were recorded during the runs; the information was used to compare apparent carriage position as determined by the radar outputs with actual carriage position as determined from the simulator outputs. Figure 8, a pictorial diagram of the test setup, shows some of the recorded parameters. Table III defines each measured parameter and indicates the source. The various velocities and

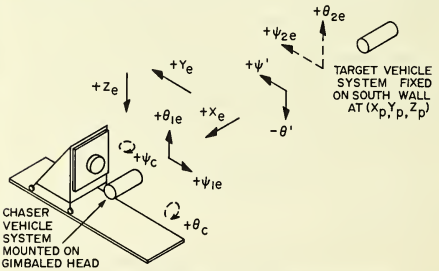


FIGURE 8. DIAGRAM OF SIMULATOR TEST SETUP.

angular rates used in the tests were representative of actual docking rates in space. The operational sequence for the dynamic runs is as follows.

a. The two vehicles are initially displaced longitudinally 17.47 meters (57.3 ft) and the chaser is either drifting or rotating into the acquisition beam from the target. Both units are searching (Phase 1).

TABLE III. DEFINITION OF RECORDED ITEMS FOR SIMULATOR TESTS.

Parameter	Source	Definition
R	Radar	Range
$\dot{R}$	Radar	Range rate
θ_{1e}	Radar	Yaw measurement from chaser system
ψ_{1e}	Radar	Pitch measurement from chaser system
θ_{2e}	Radar	Yaw measurement from target system
ψ_{2e}	Radar	Pitch measurement from target system
P	Radar	Phase signal (3 states) Phase I = 10 volts - both units searching Phase II = 5 volts - chaser detects target beacon Phase III = 0 volts - both units aligned and tracking
X_e	Simulator	Longitudinal position of carriage
Y_e	Simulator	Horizontal position of carriage
Z_e	Simulator	Vertical position of carriage
X_c	Simulator	Commanded longitudinal carriage position
θ_R	Simulator	Yaw gimbal readout
ψ_R	Simulator	Pitch gimbal readout
θ'	Simulator	Calculated yaw position from X_e, Y_e, Z_e
ψ'	Simulator	Calculated pitch position from X_e, Y_e, Z_e

b. The chaser detects target beacon (Phase 2 starts) and the drift rates are removed.

c. The chaser rotates to illuminate the target with its transmitters.

d. The target acquires the chaser beam (Phase 3 starts).

e. The chaser translates horizontally and vertically to simulate target rotation since the target is fixed. Simultaneously the chaser translates longitudinally toward the target at a rate of 0.15 m/s.

f. When the chaser is within 2.95 meters of the target, the longitudinal rate drops to 0.075 m/s.

g. When the chaser is within 2.0 meters (in the proximity of the target ring), the problem is stopped and docking is considered to be complete. Figure 9 shows the location of the vehicles at the completion of the problem.

FIGURE 9. SIMULATOR-RADAR CONFIGURATION AFTER DOCKING.

Long Range Tests

Long range tests from mountain top to aircraft were performed to test the 120 kilometer acquisition capability and to verify the long range capability. The purpose of operating at high altitude was to get above the atmospheric inversion layer, thereby eliminating a significant amount of the atmospheric disturbances.

The tests were performed with the target vehicle system on the aircraft and the chaser located on the mountain top on a tracking mount. The tracking mount served to simulate angular motion of the chaser spacecraft. Figure 10 shows the chaser vehicle system on the tracking mount.

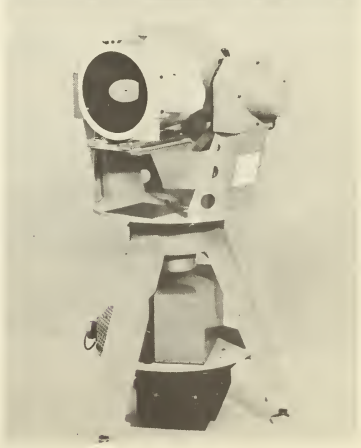


FIGURE 10. CHASER VEHICLE SYSTEM ON THE TRACKING MOUNT.

The target system was flown to various ranges; then the two systems were aligned to within the acquisition angle. Once acquisition occurred, the aircraft was flown toward the mountain top and ranging and tracking were performed until the aircraft configuration blocked the beam and caused dropout.

Test Results and Interpretation of Data

Short Range Results

The initial static range tests were found to result in large errors as the chaser system and the target system were brought together. The difficulty was discovered to be caused by multiple internal reflections of off-axis rays in the received signal. The multiple reflections caused an increase in the apparent optical path length, thereby causing the range readings to be high. This phenomena occurred only at ranges less than 10 meters.

The telescope was modified to include baffles to block off-axis rays. This greatly increased performance although the effect is still noticeable; further design can completely eliminate it.

In the first type of static tests, measurements were made over a distance of 22 meters. Typical results show rms deviation from actual range to be less than 10 centimeters. Figure 11 shows a plot of a typical run. The mean range error was found to be 5 centimeters and rms deviations from the mean were ±4.43 centimeters.

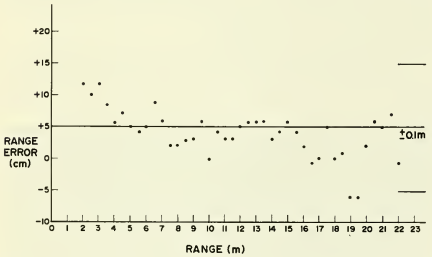


FIGURE 11. TYPICAL SHORT RANGE TEST RESULTS.

In the second type of tests, relative range errors over one meter increments were studied. Figure 12 indicates a typical run. Measurements were made every centimeter from 7 to 8 meters. Accuracy of the measurements was ±0.25 centimeters. The mean range error for the run was 3.5 centimeters and the rms deviation from the mean was only 0.85 centimeters.

The results of these tests indicate an entirely adequate performance for docking, with accuracies well within the design specifications.

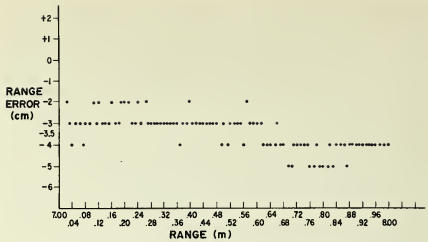


FIGURE 12. TYPICAL RELATIVE SHORT RANGE TEST RESULTS.

Rendezvous Closure and Docking Simulator Tests

The results of static tests on the simulator were not recorded because they were performed only to verify the laboratory static tests and to ensure that the system was operational.

Dynamic tests were run through the month of April 1967. During this time, several hundred runs were made, and each run culminated in a successful docking. The angular accuracy of the radar was found to exceed that which could be measured by the simulator. Test results indicate an accuracy of at least 100 arc seconds and probably more. Range and range rate measurements were well within the design specifications given in Table I.

During these runs, several areas were noted where modifications would improve the performance of the radar. Many of these were the normal engineering bugs encountered in any developmental system; all were of a minor nature and did not impede the successful operation of the radar. The recorded data of a typical run are shown in Figure 13. The channels showing angular readouts of the radar reflect a minor difficulty; i.e., the radar would not stay in the track mode until both vehicles were on-axis. The difficulty was caused by the angular dropout of the tracking systems on both vehicles. The angular dropout of the target was attributed to the lack of automatic gain control in the processing circuitry and to a poorly focused telescope. The dropout of the chaser was the result of an unsynchronized clock driving the target beacon. These problems were subsequently eliminated, and a second set of tests run early in June 1967 reflected near perfect performance. A detailed report will reflect the final results.

A plot of the final docking positions for several typical runs during the April tests is shown in Figure 14. The maximum error was less than 7.5 centimeters which is well within the requirements for docking. The modifications performed prior to the June tests should significantly improve this number.

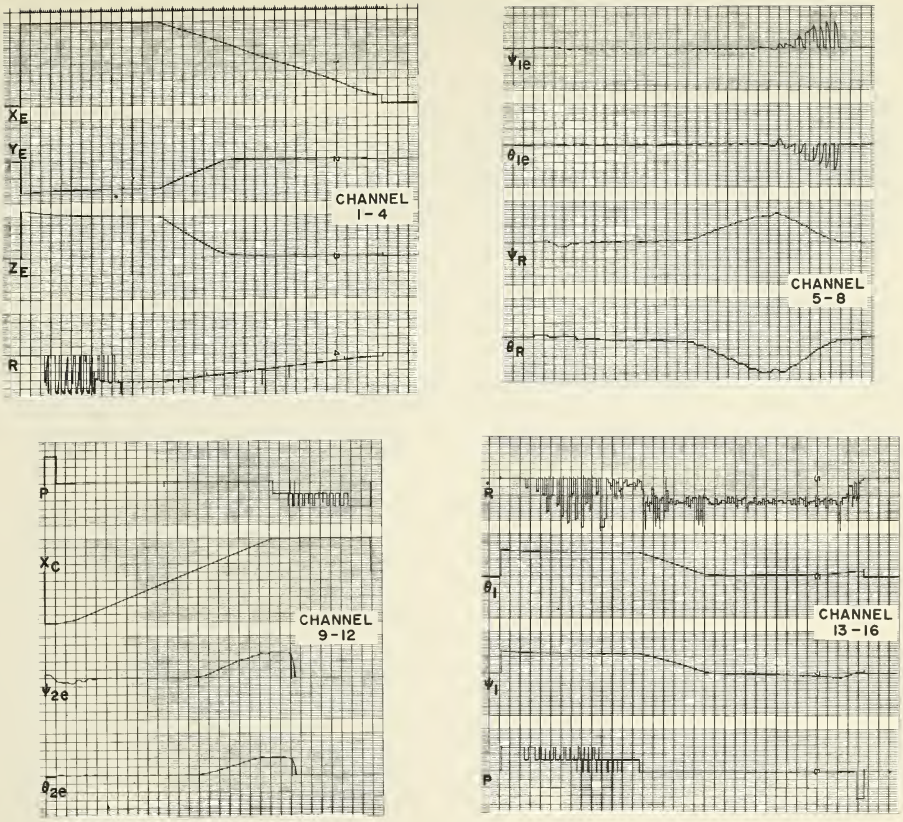


FIGURE 13. TYPICAL RECORDED DATA FOR SIMULATOR TESTS.

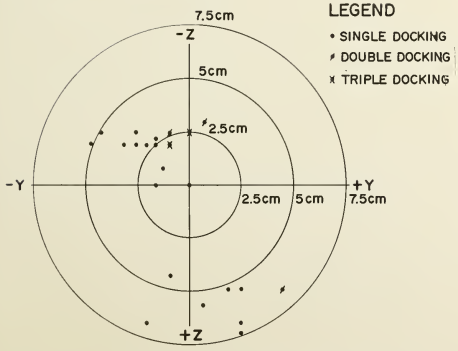


FIGURE 14. PLOT OF FINAL DOCKING POSITIONS FOR SEVERAL TYPICAL TEST RUNS.

Conclusions

The test program described here was designed to fully evaluate a radically new type of rendezvous and docking sensor. The successful development of a laser radar for rendezvous and docking represents a significant advance in the state of the art; it is one result of an intensive program at the George C. Marshall Space Flight Center to exploit the rapidly advancing technologies of lasers and optics. This program is sponsored by the Office of Advanced Research and Technology; research and development of devices, components, and techniques as well as development of systems are included in this program.

Several improvements and modifications to the radar resulted from early tests. The final test results indicate excellent performance of the radar and show that it is more than adequate for any cooperative rendezvous situation, manned or unmanned. All of the original design

specifications were met or exceeded. The tests indicate that considerably improved range and angle accuracies can be achieved if required for particular missions.

Rendezvous and docking will be necessary for future space missions, including the moon landing, orbital refueling, supply shuttles to large space stations, assembly of large complexes in space, and space rescue operations. Some missions are manned and some are unmanned; many require a completely automated rendezvous and docking capability. Equipment of the type described here can provide the guidance information for any or all of these missions.

This prototype laser radar has clearly indicated the size, weight, power, and performance advantages of a laser system over more conventional approaches. A second model of this radar will certainly exceed the prototype in every respect. Furthermore, future models will become even smaller and more versatile as improved sources and detectors become available and as more advanced optical techniques are exploited.[5]

References

1. Thomas, E. R.; Mullendore, D. L.; and Paul, V. J.; Optical Guidance Feasibility Study, Summary Report No. TP-2-83-1469-S1, Data Corp., Dayton, Ohio.
2. Dixon, Thomas P.; Wyman, Charles L.; and Coombes, H. Dean; A Laser Guidance System for Rendezvous and Docking, Navigation, Journal of the Institute of Navigation, Vol. 13, No. 3, Autumn 1966.
3. Wyman, Charles L.; Juergensen, Klaus; Kurtz, Robert L.; Hayes, Larry; and Gould, John M.; Precision Optical Tracking System for Advanced Launch Vehicles, NASA TM X-53076, June 25, 1964.
4. Smith, G. H.; and DeRocher, W. L., Jr.; A Full Scale Six-Degree-of-Freedom Orbital Closure and Docking Simulator, Martin Company, M-63-134, Denver, Colo.
5. Wyman, Charles L.; The Optical Guidance System for Rendezvous, Proceedings of the Space Optical Technology Conference, Vol. 1, Marshall Space Flight Center, Huntsville, Ala.

-- NOTES --

-- NOTES --

No. 67-618

A STUDY OF LOW-THRUST GUIDANCE

by

G. R. ASH
California Institute of Technology
Pasadena, California
and
B. M. DOBROTIN
Jet Propulsion Laboratory
Pasadena, California

AIAA Paper
No. 67-618

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

2.06, 8.07, 10.06

-- NOTES --

A STUDY OF LOW-THRUST GUIDANCE*

G. R. Ash
NASA Fellow
California Institute of Technology
Pasadena, California

B. M. Dobrotin
Group Leader, Advanced Studies Group
Guidance and Control Division
Jet Propulsion Laboratory
Pasadena, California

Abstract

The problem of guiding a low-thrust spacecraft back to the nominal trajectory has been formulated. Two closed-loop solutions have been found using a minimum-time criterion. Experimental results for a typical spacecraft with associated initial errors are presented for both solutions.

I. Introduction

In the past few years, much interest has been developed in the use of ion propulsion for space missions. The low-thrust ion engine will probably find its most important application in missions to the outer planets, where the retarding effect of the sun's gravity will require a large space vehicle energy. Up to the present, all the energy (velocity) of a spacecraft has been provided by the launch vehicle. For high-energy missions, such as those to the outer planets, it seems desirable to use high-impulse, low-thrust engines to augment the energy supplied by the boost vehicle. These low-thrust devices would operate during the long flight time between launch and encounter, supplying a higher specific impulse than that available from the present chemical boosters.

For any space mission, a nominal or desired trajectory is determined. When an ion engine is used, a nominal thrust program will be specified. The nominal trajectory selected represents the "best" possible choice in terms of many conflicting factors such as launch energy required, arrival date, telemetering and tracking considerations, ion-engine fuel required, etc.

The notion here of "best" is not really a mathematical one, since, as more factors are considered, it becomes increasingly difficult, if not impossible, to write a mathematical performance criterion that meaningfully compares one trajectory with another. Hence, the concept of neighboring or alternative nominal trajectories does not exist in this context. In addition, after a trajectory is chosen, many of the spacecraft systems are designed assuming that the vehicle will be on, or near, the nominal orbit.

However, owing to launch energy dispersion and other random effects, a spacecraft will inevitably be perturbed from its standard path. Considering the discussion above, the only alternative is to return the vehicle to the nominal trajectory. We are led therefore to inquire how an ion engine could be employed to achieve this purpose--that is, to serve the additional function of guiding the vehicle back onto the nominal trajectory.

To answer this query, the following control problems were considered:

(1) What control should be made available; i.e., how should we be able to modify the thrust vector to achieve spacecraft guidance?

(2) What criterion function should be used to judge each possible guidance system?

(3) What is the control law that minimizes the criterion in (2) subject to the control constraints in (1)?

II. Assumptions

To solve the problems posed above, the following assumptions are made:

(1) The space vehicle is in heliocentric flight. This assumption is based on the fact that the ion engine would be turned on three days after launch[1], and hence the vehicle is out of the earth's gravitational field.

(2) Spacecraft motion is constrained to one plane. This is considered a valid initial assumption and a desirable characteristic of space trajectories.

(3) The nominal thrust program consists of a constant thrust vector always at a right angle[1] to the sun-vehicle line. This is shown in Fig. 1.

(4) The acceleration level is a constant over periods of time necessary for control. That is, we are assuming the vehicle mass is a constant over the period of interest.

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

67-618

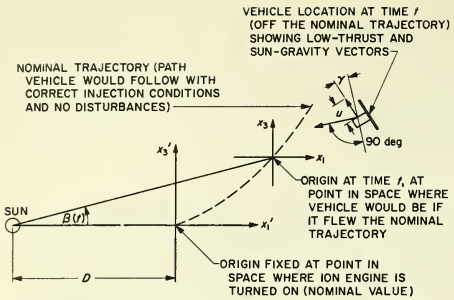


Figure 1. Definition of the coordinate frames (x_1', x_3') and (x_1, x_3)

(5) The state vector of the vehicle (i.e., position and velocity) is known at all times. This information is available from orbit determination via tracking data.

If control in two dimensions is to be obtained, one feels that it would be advantageous to have independent controls available in two directions. One practical way of obtaining such control would be, first, to allow small attitude variations and, second, to allow the acceleration level to change slightly. The control scheme used in this study allows only nine discrete states of the thrust vector, including the nominal state (see Fig. 2). This implementation has the advantage of being simple and highly realistic.

The criterion we will use will be minimum time--that is, we wish to guide the vehicle back onto the nominal trajectory in as little time as possible. This is felt to be a good criterion in that it is desired to have the spacecraft on the nominal orbit for as much of the flight as possible. Also, it is known that the solution of the minimum time problem involves "bang-bang" control, or using discrete levels of control. Since we have constrained our thrust vector to be discrete, this configuration seems well suited for a minimum time controller.

To solve for the control law, the theory of optimal control was applied. Actually, two solutions to this problem were obtained, and they will be discussed below.

III. First Solution

The coordinate systems we will be considering appear in Fig. 1. The coordinate frame (x_1, x_3) has its origin at time t at the point in space where a vehicle on the nominal trajectory would be at time t, assuming flight begins at t = 0. Note that the angle $\beta(t)$ is determined by the nominal trajectory desired and is thus a function of time only. The differential equations of vehicle motion

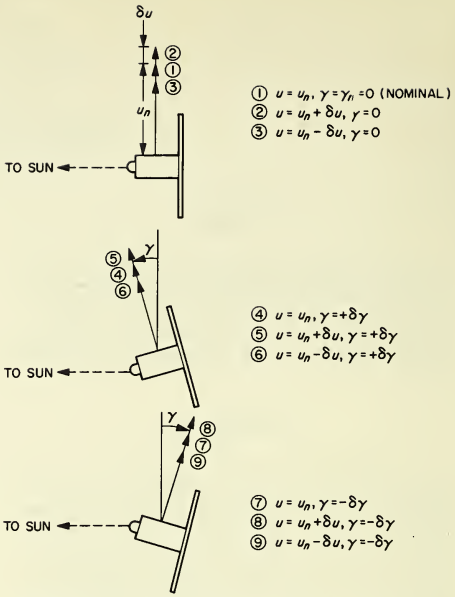


Figure 2. The nine allowable states of the ion-engine thrust vector

in the (x_1', x_3') frame are as follows (dots indicate derivatives with respect to time):

$$\dot{x}_1' = x_2' \triangleq F_1$$

$$\dot{x}_2' = \ddot{x}_1' = \frac{-GM_s(x_1' + D)}{\left((x_1' + D)^2 + (x_3')^2\right)^{3/2}} - \frac{u\left(x_3' \cos\gamma + (x_1' + D)\sin\gamma\right)}{\left((x_1' + D)^2 + (x_3')^2\right)^{1/2}} \triangleq F_2$$

$$\dot{x}_3' = x_4' \triangleq F_3$$

$$\dot{x}_4' = \ddot{x}_3' = \frac{-GM_s(x_3')}{\left((x_1' + D)^2 + (x_3')^2\right)^{3/2}} + \frac{u\left((x_1' + D)\cos\gamma - x_3'\sin\gamma\right)}{\left((x_1' + D)^2 + (x_3')^2\right)^{1/2}} \triangleq F_4 \qquad (1)$$

where G is the constant of gravitation, M_s is the mass of the sun, and u, γ, and D are defined in Figs. 1 and 2.

In vector notation, Eq. (1) becomes

$$\dot{X}' = F(u, \gamma, X')$$

where $X' = \text{col}[x_1', x_2', x_3', x_4']$ and $F = \text{col}[F_1, F_2, F_3, F_4]$. If at time τ we have deviations $X(\tau) = \text{col}[x_1(\tau), x_2(\tau), x_3(\tau), x_4(\tau)]$ from the nominal trajectory, then the problem is to find the controls

$$u(t), \quad \gamma(t) \quad 0 \leq \tau \leq t \leq T$$

such that

$$X(T) = 0$$

and the performance index

$$\int_\tau^T dt$$

is minimized.

Referring to Fig. 3, consider the following coordinate transformation:

$$\left.\begin{aligned} x_1'' &= x_1' \cos\beta + x_3' \sin\beta \\ x_3'' &= -x_1' \sin\beta + x_3' \cos\beta \end{aligned}\right\} \tag{2}$$

Differentiating these equations twice yields

$$\left.\begin{aligned} \dot{x}_1'' &= \dot{x}_1' \cos\beta + \dot{x}_3' \sin\beta - x_1'\dot{\beta}\sin\beta + x_3'\dot{\beta}\cos\beta \triangleq x_2'' \\ \ddot{x}_1'' &= \ddot{x}_1' \cos\beta + \ddot{x}_3' \sin\beta - 2\dot{x}_1'\dot{\beta}\sin\beta - x_1'\dot{\beta}^2\cos\beta - x_1'\ddot{\beta}\sin\beta + 2\dot{x}_3'\dot{\beta}\cos\beta - x_3'\dot{\beta}^2\sin\beta + x_3'\ddot{\beta}\cos\beta \\ \dot{x}_3'' &= -\dot{x}_1' \sin\beta + \dot{x}_3' \cos\beta - x_1'\dot{\beta}\cos\beta - x_3'\dot{\beta}\sin\beta \triangleq x_4'' \\ \ddot{x}_3'' &= \dot{x}_4'' = -\ddot{x}_1' \sin\beta + \ddot{x}_3' \cos\beta - 2\dot{x}_1'\dot{\beta}\cos\beta + x_1'\dot{\beta}^2\sin\beta - x_1'\ddot{\beta}\cos\beta - 2\dot{x}_3'\dot{\beta}\sin\beta - x_3'\dot{\beta}\cos\beta - x_3'\ddot{\beta}\sin\beta \end{aligned}\right\} \tag{3}$$

Letting the subscript n denote the nominal values of the variables $\dot{X}''$, u, γ, and X', Eq. (3) can be written

$$\dot{X}_n'' + \delta\dot{X}'' = R(t)F(u_n + \delta u, \gamma_n + \delta\gamma, X_n' + X) + S(t)(X_n' + X) \tag{4}$$

where $\delta X''$, δu, and $\delta\gamma$ are deviations from nominal values, and the matrices R(t) and S(t) are defined by

$$R(t) \triangleq \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & \cos\beta & 0 & \sin\beta \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & -\sin\beta & 0 & \cos\beta \end{bmatrix} \tag{5a}$$

$$S(t) \triangleq \begin{bmatrix} -\dot{\beta}\sin\beta & 0 & \dot{\beta}\cos\beta & 0 \\ -\ddot{\beta}\sin\beta - \dot{\beta}^2\cos\beta & -2\dot{\beta}\sin\beta & \ddot{\beta}\cos\beta - \dot{\beta}^2\sin\beta & 2\dot{\beta}\cos\beta \\ -\dot{\beta}\cos\beta & 0 & -\dot{\beta}\sin\beta & 0 \\ -\ddot{\beta}\cos\beta + \dot{\beta}^2\sin\beta & -2\dot{\beta}\cos\beta & -\ddot{\beta}\sin\beta - \dot{\beta}^2\cos\beta & -2\dot{\beta}\sin\beta \end{bmatrix} \tag{5b}$$

In space trajectory problems such as this one, linearization of the dynamical equations about a nominal path often yields satisfactory approximations to the true physical situation. This statement is supported by the fact that expected injection errors and disturbances en route, as well as the quantities δu and $\delta\gamma$, are "small" in relation to their nominal values. Hence, we carry out a Taylor series expansion of Eq. (4), which yields

$$\dot{X}'_n + \delta\dot{X}' = R(t)\Big(F(u_n, \gamma_n, X'_n) + F_{X'}(u_n, \gamma_n, X'_n)X + F_u(u_n, \gamma_n, X'_n)\delta u + F_\gamma(u_n, \gamma_n, X'_n)\delta\gamma + (\text{higher-order terms})\Big) + S(t)(X'_n + X) \tag{6}$$

where $F_{X'}$ is the Jacobian matrix whose ij-th element, $\partial F_i/\partial x'_j$, is the partial derivative of the ith component of F with respect to the jth component of X'. Also, $F_u = \text{col}[\partial F_1/\partial u, \partial F_2/\partial u, \partial F_3/\partial u, \partial F_4/\partial u]$ and similarly for F_γ. Defining the controls

$$u_1 = -u_n\delta\gamma \qquad u_2 = \delta u \tag{7}$$

and making suitable manipulations, Eq. (6) becomes

$$\delta\dot{X}'(t) = \begin{bmatrix} 0 & \cos\beta(t) & 0 & \sin\beta(t) \\ A\cos\beta(t) + C\sin\beta(t) & 0 & B\cos\beta(t) + D\sin\beta(t) & 0 \\ 0 & -\sin\beta & 0 & \cos\beta \\ -A\sin\beta(t) + C\cos\beta(t) & 0 & -B\sin\beta(t) + D\cos\beta(t) & 0 \end{bmatrix} X(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u_2(t)$$

$$+ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u_1(t) + \begin{bmatrix} -\dot{\beta}\sin\beta & 0 & \dot{\beta}\cos\beta & 0 \\ -\ddot{\beta}\sin\beta - \dot{\beta}^2\cos\beta & -2\dot{\beta}\sin\beta & \ddot{\beta}\cos\beta - \dot{\beta}^2\sin\beta & 2\dot{\beta}\cos\beta \\ -\dot{\beta}\cos\beta & 0 & -\dot{\beta}\sin\beta & 0 \\ -\ddot{\beta}\cos\beta + \dot{\beta}^2\sin\beta & -2\dot{\beta}\cos\beta & -\ddot{\beta}\sin\beta - \dot{\beta}^2\cos\beta & -2\dot{\beta}\sin\beta \end{bmatrix} X(t) \tag{8}$$

where higher-order terms have been neglected. Here

$$A = \frac{-GM_s + u_n(x'_{1n} + D)}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{3/2}} + \frac{3GM_s(x'_{1n} + D)^2}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{5/2}}$$

$$B = \frac{-u_n(x'_{1n} + D)^2}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{3/2}} + \frac{3GM_s(x'_{1n} + D)x'_{3n}}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{3/2}}$$

$$C = \frac{3GM_s x'_{3n}(x'_{1n} + D)}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{5/2}} + \frac{u_n\left(2(x'_{1n} + D)^2 + (x'_{3n})^2\right)}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{3/2}}$$

$$D = \frac{GM_s\left(2(x'_{3n})^2 - (x'_{1n} + D)^2\right)}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{5/2}} - \frac{u_n x'_{3n}(x'_{1n} + D)}{\left((x'_{1n} + D)^2 + (x'_{3n})^2\right)^{3/2}}$$

To gain more insight into the problem, Eq. (8) will be simplified by neglecting small terms. The quantities A, B, C, and D are proportional to changes in the sun's gravity and the angle β over a region in space near the nominal trajectory. These quantities are of the order of 10^{-12}, in mks units, and will therefore be neglected. For the Mars mission under consideration, the quantities $\ddot{\beta}$ and $\dot{\beta}^2$ are of the order 10^{-14} and will also be neglected. Finally, it will be assumed that the quantities $\dot{\beta}x_2$ and $\dot{\beta}x_4$ are negligible with respect to u_1 and u_2. Actually, typical values would be 10^{-6} for $\dot{\beta}x_2$ and $\dot{\beta}x_4$, and 10^{-4} for u_1 and u_2. We expect that the latter assumption will give by far the largest error.

We use the physical reasoning above and the fact that

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \delta X''(t) = \begin{bmatrix} -\dot{\beta}\sin\beta & \cos\beta & \dot{\beta}\cos\beta & \sin\beta \\ 0 & 0 & 0 & 0 \\ -\dot{\beta}\cos\beta & -\sin\beta & -\dot{\beta}\sin\beta & \cos\beta \\ 0 & 0 & 0 & 0 \end{bmatrix} X(t)$$

to obtain

$$\delta\dot{X}''(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \delta X''(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u_1(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} u_2(t) \qquad (9)$$

It is evident from examining Eq. (9) that the four-dimensional minimum-time problem has been reduced to two two-dimensional problems, since the $\delta x''_1$ and $\delta x''_2$ equations are decoupled from the $\delta x''_3$ and $\delta x''_4$ equations. The minimum-time problem may now be easily solved by applying the basic principles of optimal control theory. We will solve one two-dimensional problem.

Let

$$\delta x''_1 = y_1 \qquad \delta x''_2 = y_2$$

then

$$\dot{y}_1 = y_2 \qquad \dot{y}_2 = u_1$$

The Lagrange multiplier equations for this problem are

$$\dot{\lambda}_1 = 0 \qquad \dot{\lambda}_2 = -\lambda_1$$

whose solutions are

$$\lambda_1(t) = \lambda_1(0) \qquad \lambda_2(t) = -\lambda_1(0)t + \lambda_2(0) \qquad (11)$$

The optimal control is

$$u_1 = -\text{sgn}(\lambda_2)$$

and, from Eq. (11), only one "switching" of the control u_1 is possible. Solving Eq. (10) for constant u_1 yields

$$y_1(t) = \tfrac{1}{2}u_1 t^2 + y_2(0)t + y_1(0)$$

$$y_2(t) = u_1 t + y_2(0)$$

Eliminating t from these equations, we find that

$$2u_1\left(y_1 - y_1(0)\right) = \left(y_2 - y_2(0)\right)^2 + 2y_2(0)\left(y_2 - y_2(0)\right) \tag{12}$$

Equation (12) shows that the vehicle will follow a parabolic trajectory in the (y_1, y_2) plane for constant u_1. Coupling this fact with the fact that only one switching is optimal, the "switching boundary" is obtained, as shown in Fig. 4. The expected trajectory for a set of initial deviations from the nominal trajectory is also shown in Fig. 4.

IV. Experimental Results of the First Solution

One practical and meaningful way to check the results above is by digital computer simulation. To that end, an initial error of about 12 m/sec was assumed and is considered typical[1]. An initial position error that would result from 3 days of velocity error propagation was used. Vehicle weight was taken as 4,535 kg, nominal low-thrust acceleration as 10^{-3} m/sec, and u_1 and u_2 as 10^{-4} m/sec^2.

It was found that the error incurred by neglecting $\dot{\beta}x_1$ and $\dot{\beta}x_3$ was large enough to require that the minimum-time solution be applied twice; that is, the large initial errors were reduced, and then the resulting errors were similarly reduced. The trajectories obtained are shown in Fig. 5.

Probably the most significant disturbance on a practical system will be the attitude-control limit cycle operation causing attitude variation of the thrust vector. A sinusoidal disturbance with an amplitude of 0.5 deg and a frequency of 1 cycle per 20 min was put into the control system. The resulting trajectories are shown in Fig. 6, and the performance is seen to be good.

V. Second Solution

The motivation for the second solution is the desirability of obtaining a more accurate approximation to the minimum-time solution by making larger use of digital computer capabilities. It is desirable to avoid numerical solution of differential equations, however, which is costly in computer time.

To begin with, we will linearize Eq. (1) as follows:

$$\dot{X}'(t) = \dot{X}'_n(t) + \dot{X}(t) = F(u_n + \delta u, \gamma_n + \delta\gamma, X'_n + X) = F(u_n, \gamma_n, X'_n) + F_u\delta_u + F_\gamma\delta_\gamma + F_{X'}X + (\text{higher-order terms})$$

As before, we neglect higher-order terms, and the terms A, B, C, and D in $F_{X'}$. Also, we use the definitions of u_1 and u_2 to obtain

$$\dot{X}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} X(t) + \begin{bmatrix} 0 \\ \cos\beta(t) \\ 0 \\ \sin\beta(t) \end{bmatrix} u_1(t) + \begin{bmatrix} 0 \\ -\sin\beta(t) \\ 0 \\ \cos\beta(t) \end{bmatrix} u_2(t) \tag{13}$$

The differential equations for the Lagrange multipliers are then

$$\dot{\lambda}_1 = 0 \qquad \dot{\lambda}_2 = -\lambda_1 \qquad \dot{\lambda}_3 = 0 \qquad \dot{\lambda}_4 = -\lambda_3$$

which have the solutions

$$\lambda_1(t) = \lambda_1(0) \qquad \lambda_2(t) = -\lambda_1(0)t + \lambda_2(0) \qquad \lambda_3(t) = \lambda_3(0) \qquad \lambda_4(t) = -\lambda_3(0)t + \lambda_4(0)$$

The optimal controls are therefore given by

$$\left.\begin{aligned} u_1^*(t) &= -\mathrm{sgn}\Big(\big(-\lambda_1(0)t + \lambda_2(0)\big)\cos\beta(t) + \big(-\lambda_3(0)t + \lambda_4(0)\big)\sin\beta(t)\Big) \\ u_2^*(t) &= -\mathrm{sgn}\Big(\big(-\lambda_1(0)t + \lambda_2(0)\big)\big(-\sin\beta(t)\big) + \big(-\lambda_3(0)t + \lambda_4(0)\big)\cos\beta(t)\Big) \end{aligned}\right\} \tag{14}$$

Some possible realizations of Eq. (14) would be as in Fig. 7. (Note that β is not expected to exceed 90 deg before nominal trajectory acquisition.) These realizations suggest that each control would have a maximum of two switchings.

Now, given the initial conditions on Eq. (12), we can write the explicit solution for X(T), where T is the nominal trajectory acquisition time. That is,

$$X(T) = \phi(T,0)X(0) + \int_0^T \phi(T,t)\begin{bmatrix} 0 \\ \cos\beta(t) \\ 0 \\ \sin\beta(t) \end{bmatrix} u_1(t)\,dt + \int_0^T \phi(T,t)\begin{bmatrix} 0 \\ -\sin\beta(t) \\ 0 \\ \cos\beta(t) \end{bmatrix} u_2(t)\,dt \tag{15}$$

where $\phi(t_2, t_1)$ is the fundamental matrix that satisfies the matrix differential equation

$$\dot{\phi}(t_2,t_1) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\phi(t_2,t_1) \tag{16}$$

with $\phi(t_1, t_1) = I$. The solution of Eq. (16) is

$$\phi(t_2,t_1) = \begin{bmatrix} 1 & (t_2 - t_1) & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & (t_2 - t_1) \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17}$$

Since the absolute values of u_1 and u_2 are constant, only the sign of these quantities is needed inside the integrals of Eq. (15). If we designate $u_1(0)$ and $u_2(0)$ as the initial values of u_1 and u_2, t_1 and t_2 as the switching times of u_1, and t_3 and t_4 as the switching times of u_2, then Eq. (15) becomes

$$X(T) = \phi(T,0)X(0) + u_1(0)\left(\int_0^{t_1} - \int_{t_1}^{t_2} + \int_{t_2}^{T}\right) + u_2(0)\left(\int_0^{t_3} - \int_{t_3}^{t_4} + \int_{t_4}^{T}\right) \tag{18}$$

The integrals of Eq. (18) can be explicitly evaluated if we assume that β varies at a constant rate. This is an excellent approximation for the trajectories of interest. Hence, if we assume that

$$\beta(t) = \omega t \qquad \omega = \text{constant} \doteqdot \dot{\beta}$$

and if we define

$$I_1 = u_1(0)\left(\frac{2t_1}{\omega}\sin\omega t_1 + \frac{2}{\omega^2}\cos\omega t_1 - \frac{2t_2}{\omega}\sin\omega t_2 - \frac{2\cos\omega t_2}{\omega^2} + \frac{T\sin\omega T}{\omega} + \frac{1}{\omega^2}\cos\omega T - \frac{1}{\omega^2}\right)$$

$$- u_2(0)\left(-\frac{2t_3\cos\omega t_3}{\omega} + \frac{2\sin\omega t_3}{\omega^2} + \frac{2t_4\cos\omega t_4}{\omega} - \frac{2\sin\omega t_4}{\omega^2} - \frac{T\cos\omega T}{\omega} + \frac{\sin\omega T}{\omega^2}\right)$$

$$I_2 = u_1(0)\left(\frac{2}{\omega}\sin\omega t_1 - \frac{2}{\omega}\sin\omega t_2 + \frac{1}{\omega}\sin\omega T\right) - u_2(0)\left(-\frac{2}{\omega}\cos\omega t_3 + \frac{2}{\omega}\cos\omega t_4 - \frac{1}{\omega}\cos\omega T + \frac{1}{\omega}\right)$$

$$I_3 = u_1(0)\left(\frac{-2t_1\cos\omega t_1}{\omega} + \frac{2\sin\omega t_1}{\omega^2} + \frac{2t_2\cos\omega t_2}{\omega} - \frac{2\sin\omega t_2}{\omega^2} - \frac{T\cos\omega T}{\omega} + \frac{\sin\omega T}{\omega^2}\right)$$

$$+ u_2(0)\left(\frac{2t_3}{\omega}\sin\omega t_3 + \frac{2}{\omega^2}\sin\omega t_3 - \frac{2t_4}{\omega}\sin\omega t_4 - \frac{2\cos\omega t_4}{\omega^2} + \frac{T\sin\omega T}{\omega} + \frac{1}{\omega^2}\cos\omega T - \frac{1}{\omega^2}\right)$$

$$I_4 = u_1(0)\left(\frac{-2\cos\omega t_1}{\omega} + \frac{2}{\omega}\cos\omega t_2 - \frac{1}{\omega}\cos\omega T + \frac{1}{\omega}\right) + u_2(0)\left(\frac{2}{\omega}\sin\omega t_3 - \frac{2}{\omega}\sin\omega t_4 + \frac{1}{\omega}\sin\omega T\right)$$

then Eq. (18) becomes

$$\left.\begin{aligned} x_1(T) &= x_1(0) + Tx_2(0) - I_1 + TI_2 \triangleq G_1(t_1, t_2, t_3, t_4, T) \\ x_2(T) &= x_2(0) + I_2 \triangleq G_2(t_1, t_2, t_3, t_4, T) \\ x_3(T) &= x_3(0) + Tx_4(0) - I_3 + TI_4 \triangleq G_3(t_1, t_2, t_3, t_4, T) \\ x_4(T) &= x_4(0) + I_4 \triangleq G_4(t_1, t_2, t_3, t_4, T) \end{aligned}\right\} \quad (19)$$

Equations (19) are four equations in five unknowns. Since it is desired that X(T) = 0, the problem is now to find the minimum value of T for which Eqs. (19) can be satisfied. Fortunately enough, these equations can be solved by the Newton-Raphson technique, and such analysis indicates that the minimum value of T is achieved either when $T = t_2$ or $T = t_4$. Hence one control will have one switching, and the other will have two switchings. It is fairly easy to determine the correct $u_1(0)$ and $u_2(0)$, and thereby Eqs. (19) can be solved for the minimum value of T and for the switching times of the control variables.

V. Experimental Results of the Second Solution

The computer simulation of the solution obtained above indicated that there was a need to account for second-order effects in the control variables. Hence, for the case when $T = t_4$, Eq. (18) can be modified as follows:

$$X(T) = \phi(t, 0)X(0) + u_1(0)\left(\text{FAC 1}\int_0^{t_1} - \text{FAC 1}\int_{t_1}^{t_3} - \text{FAC 2}\int_{t_3}^{t_2} + \text{FAC 2}\int_{t_2}^{T}\right)$$

$$+ u_2(0)\left(\text{FAC 3}\int_0^{t_3} - \text{FAC 4}\int_{t_3}^{T}\right)$$

where FAC 1, FAC 2, FAC 3, and FAC 4 are the factors that account for the second-order effects. As a result of integrating these equations, one obtains answers very similar to Eqs. (19), and a solution can again be obtained using the Newton-Raphson technique.

Excellent performance was obtained using this modified solution. The results appear in Fig. 8, and a comparison with Fig. 5 indicates that the second solution (1) requires about 24 hours less time to acquire the nominal trajectory and (2) requires 3 fewer commands (switchings) to be sent to the vehicle for the initial deviations considered. It was also found that smaller values (i.e., 0.25×10^{-4} m/sec^2) of u_1 and u_2 could be used with no penalty. Finally, the effect of attitude variations on this solution appeared negligible with respect to the attitude control model used.

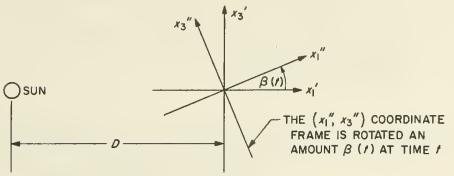


Figure 3. Definition of the (x_1', x_3') coordinate frame

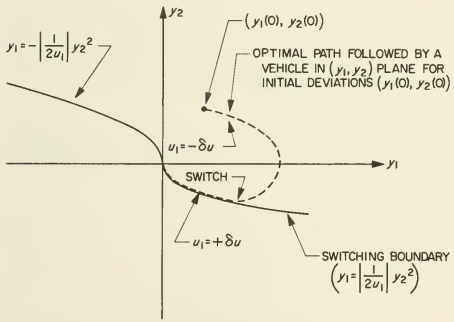


Figure 4. Definition of the "switching boundary" in the (y_1, y_2) plane

Reference

1. Solar Powered Electric Propulsion Spacecraft Study, Final Report, JPL Contract No. 951144, Hughes Aircraft Company, December 1965.

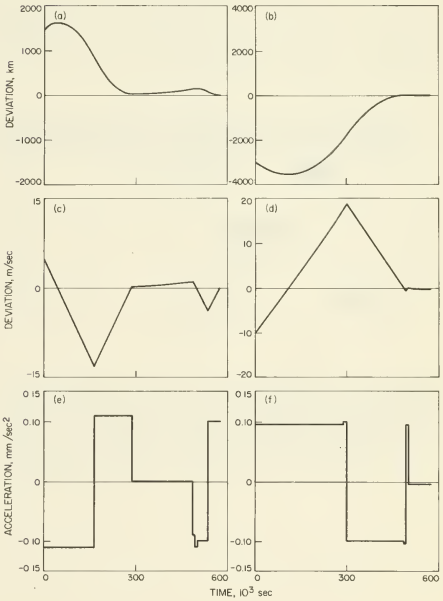


Figure 5. Results obtained for the first solution: (a) the x_1 position deviation; (b) the x_3 position deviation; (c) the x_1 velocity deviation; (d) the x_3 velocity deviation; (e) the control variable u_1; and (f) the control variable u_2

67-618

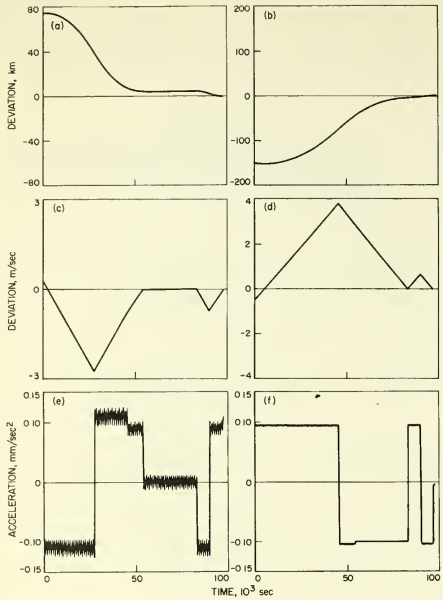


Figure 6. Results obtained for the first solution, with attitude variations; (a) the x_1 position deviation; (b) the x_3 position deviation; (c) the x_1 velocity deviation; (d) the x_3 velocity deviation; (e) the control variable u_1; and (f) the control variable u_2

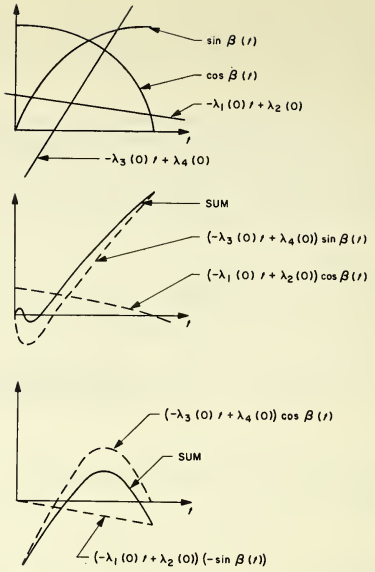


Figure 7. Control function switchings for the second solution

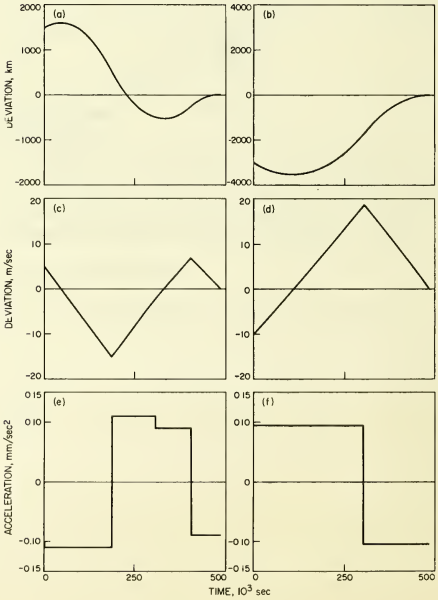


Figure 8. Results obtained for the second solution: (a) the x_1 position deviation; (b) the x_3 position deviation; (c) the x_1 velocity deviation; (d) the x_3 velocity deviation; (e) the control variable u_1; and (f) the control variable u_2

No. 67-619

PROPULSIVE-PHASE GUIDANCE FOR UNMANNED LANDING ON MARS: A TWO-PART STUDY

by

J. W. MOORE and M. J. CORK
Jet Propulsion Laboratory
Pasadena, California

AIAA Paper
No. 67-619

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.02, 8.04

-- NOTES --

PROPULSIVE-PHASE GUIDANCE FOR UNMANNED LANDING ON MARS: A TWO-PART STUDY*

J. W. Moore,** and M. J. Cork†
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Abstract

Results of a two-part study of the guidance system requirements for a Martian propulsive landing capsule are discussed. Part I considers a rough-landing system which utilizes a solid propellant retro rocket and a relatively simple guidance mechanization to achieve surface impact velocities of less than 300 fps. Ideal system performance over a range of initial conditions is determined for the assumed mechanization. Effects of radar errors on burnout altitude and impact velocity are determined from a Monte Carlo analysis. Wind effects are superimposed on the Monte Carlo results to obtain an estimate of overall system performance. Part II is an evaluation of a soft-landing system designed to provide a landing velocity not to exceed 25 fps by using throttlable, liquid-propellant engines, and a more complex mechanization. Of primary concern is the guidance system design which is complicated by the large uncertainty in the environmental conditions. Feasibility of the proposed mechanization is evaluated by using a computer simulation of the propulsive terminal descent phase. Nominal descent characteristics are determined and the effects of system perturbations caused by measurement errors and terrain slopes are considered.

Nomenclature

General Nomenclature for Part I

N = ratio for retro thrust to gravitational force on vehicle

R = slant range along flight path, ft

T = retro thrust, lbf

V = velocity, fps

VM-3, -4, -7, -8 = atmospheric models defined in Ref. 3

g_m = gravitational acceleration of Mars, 12.3 ft/sec^2

h = altitude, ft

t = time, sec

t_b = burning time of retro rocket motor, sec

x = lateral displacement, ft

γ = flight path angle, deg

$\Delta(\ )$ = increment of ( )

Δt, Δt_1 = time between two range marks, sec

$\Delta t2$ = time between second range mark and retro ignition, sec

ΔV = velocity increment provided by retro propulsion

θ = inertial orientation of thrust vector

σ = standard deviation

Subscripts

1 = first range mark

2 = second range mark

3 = retro ignition

4, bo = retro thrust termination (burnout)

Superscripts

$\cdot$ = time derivative $\left(\frac{d}{dt}\right)$

$-$ = average value

$*$ = measured value

General Nomenclature for Part II

A_T = retro-engine thrust acceleration, ft/sec^2

$C_{N,A}$ = coefficient of aerodynamic force, normal and axial

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

**Senior Engineer, Guidance and Control, Analysis and Integration Section.

†Senior Engineer, Voyager Propulsion Section.

67-619

$F_{N,A}$ = aerodynamic force, normal and axial

M = capsule mass

R = slant range, ft

V_A = aerodynamic velocity, ft/sec

V_g = surface-referenced velocity, ft/sec

V_w = wind velocity, ft/sec

$a_{N,A}$ = aerodynamic acceleration, normal and axial

g_m = gravitational acceleration of Mars, 12.3 ft/sec^2

h = altitude, ft

α = angle-of-attack, deg

α_K = acceleration contour shaping constant, ft^2/sec^2

γ_A = aerodynamic flight-path angle, deg

γ_g = surface-referenced flight-path angle, deg

μ = surface slope, deg

Subscripts

I = condition at ignition

I. Introduction

Exploration of Mars has been a prime goal of the United States Space Program since its inception. Scientific interest in this goal was reemphasized in a study, "Biology and the Exploration of Mars," conducted by the U.S. National Academy of Sciences in 1964[1]. The Jet Propulsion Laboratory (JPL), in its role as a contract agency for the National Aeronautics and Space Administration (NASA), conducted extensive studies between 1960 and 1966 concerning the placement of instrument packages on Mars[2].

Most of these studies were based on a three-stage deceleration system consisting of (1) an aeroshell for hypersonic and supersonic deceleration; (2) a parachute for transonic through low supersonic deceleration; and (3) an omnidirectional impact limiter to remove the last few hundred feet per second. A primary advantage of such a system approach is its inherent simplicity because of a relatively passive operational mode.

During the same time period adoption of Mars atmospheric models with reduced surface pressures resulted in decreased performance for an aerodynamic system. Concurrently, the scientific community developed an evolution of sophisticated payload concepts which are not compatible with the impact limiter cocoon and the 3000-g deceleration of a 250- to 300-fps impact. Because of such a limiting situation for aerodynamic decelerators, a propulsion terminal descent that requires an active guidance and control system must be considered. The aeroshell remains an effective deceleration device if the entry corridor is restricted to shallow flight-path angles, so propulsion is competitive only as a replacement for parachute and/or impact attenuator.

Two types of propulsion landing schemes have been used in the Unmanned Lunar Program. Ranger Block II used a spin-stabilized solid rocket, ignited on a radar range mark, to slow a payload with omnidirectional impact limiter to an acceptable terminal velocity of less than 200 fps. The Surveyor Program currently uses a throttled vernier propulsion system, controlled by a complex on-board guidance system, to achieve a touchdown velocity of 10 to 15 fps. These two concepts represent extremes of guidance and control mechanizations in terms of performance (terminal velocity) and complexity.

Variations on the Ranger and Surveyor mechanizations were studied in an attempt to bound the propulsion landing problem on Mars. Performance analyses were conducted to (1) compare the simple propulsion mechanization to a parachute for an early mission, and (2) evaluate the performance of the more complex mechanization for use with a large scientific payload. Because the studies were conducted at different times, for different payloads, and with slightly different ground rules, they are presented herein as two separate parts.

A parachute was not included as an intermediate stage in the soft-landing study because (1) it was desirable to minimize the total number of subsystems, and (2) the state-of-the-art of a supersonic parachute ($1 \leq$ Initial Mach No. ≤ 2) was unknown at the time of the study. As such, the system studied does not necessarily represent a preferred mode for terminal descent mechanization, but rather one possible approach. The aeroshell-parachute-propulsion scheme should be considered as an alternate concept. Although one additional subsystem is introduced, the overall system performance may show significant improvement.

II. Part I: Rough-Landing System

Mechanization and Environmental Assumptions

The Ranger Block II terminal descent system used a fixed impulse solid rocket motor which was ignited at a pre-selected range upon a signal provided by an altitude marking radar[4]. Such a simple mechanization was possible because the initial conditions of the propulsion descent phase were predictable and were relatively constant from mission to mission.

The existence of an atmosphere on Mars, and uncertainties in the density of that atmosphere, create complications which the Ranger system, in a Lunar vacuum, did not face. Specifically, the terminal velocity of the entry system varies from 500 to 1000 fps for variations in surface pressure of 5 to 10 mb. If it is assumed that a fixed-thrust solid rocket is used to provide terminal deceleration, a cursory analysis will show that both the amount of velocity increment provided by retro propulsion, ΔV, and the ignition altitude must be varied as a function of initial velocity.

If velocity must be determined with an on-board system, it must be assumed that some mechanization is necessary. The system selected consists of a computer which calculates ignition time and

burn duration based on elapsed time between two range marks. The complexity of this system is equivalent to that of a parachute sequencing system. It was chosen for the study to establish (1) a lower bound on guidance accuracy, and (2) an estimate of potential improvement with greater guidance complexity.

Figure 1 illustrates the terminal descent profile. The entry vehicle is initially descending in an aerodynamically stable mode with an altitude marking radar oriented forward along the roll axis. (An active rate stabilization system is employed to damp out vehicle oscillations during the period from range measurement to propulsion ignition.) The radar system triggers a counter at a pre-selected slant range (t_1); a second mark is registered at the next range gate (t_2), and the guidance computer calculates Δt_2 and t_b using the measured elapsed time between the two marks. These time increments are then used to initiate and terminate the propulsion phase by adding them to the reference time t_1 of the counter.

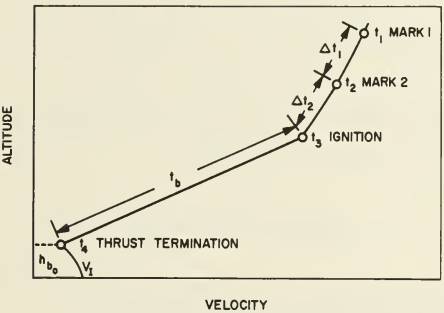


Fig. 1. Terminal descent profile

Figure 2 provides a description of the propulsive descent phase. At t_3 the descent rocket motor is ignited and the payload flies rearward from the aeroshell. The payload is assumed to remain inertially fixed during the motor burn; this could be provided by spin stabilization or an active thrust vector control system. At t_4, the thrust is terminated by nozzle jettison, the remainder of the structure is separated from the impact limiter, and the payload is allowed to fall to the surface.

Specific atmospheric, vehicle, and trajectory models were assumed to arrive at a set of initial conditions. The atmospheric models assumed were VM-3, -4, -7, -8 [3]; these models cover a range of surface pressures from 5 to 10 mb and represent two composition models. Also assumed were a ballistic coefficient (M/C_DA) of 0.35, entry flight-path angles of 15 to 20 deg below the horizontal, and entry velocity of 14,300 fps. (Entry is defined to occur at 800,000 ft.) The selection of these parameters is based on trade-offs involving landing site selection, vehicle drag, vehicle size, and terminal descent conditions [2].

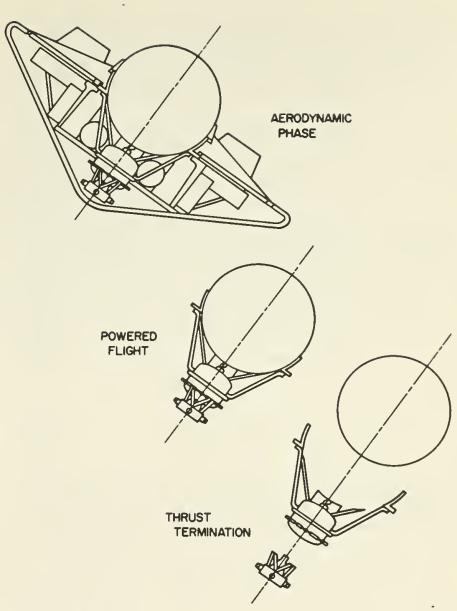


Fig. 2. Rough-landing system configuration

The terminal descent trajectory is a function of the retro propulsion thrust. A high-thrust system ignited near the surface tends to minimize propulsion weight, but the allowable thrust level is generally bounded by propulsion design constraints, i.e., number of motors, geometric constraints, and propellant burning rate. A thrust-to-Mars-weight ratio of 7.85 was selected for this study as being close to the upper limit for a single solid motor with a reasonable length-to-diameter ratio.

Analytical Model

The mechanization assumptions discussed in the previous section are expanded into an analytical model that will be used as a starting point for the guidance and control analysis. The following are the trajectory assumptions used for the derivations presented:

1. Acceleration of gravity is constant.

2. Flat planet surface.

3. Negligible atmospheric drag from ignition to impact.

4. Zero angle of attack at time of propulsion ignition.

5. Constant thrust acceleration of 3 earth-g (7.85 g_m).

6. Constant inertial angle equal to the flight-path angle at ignition.

Figure 3 shows the coordinate system and also defines some of the symbols used in the analysis. The motion of the spacecraft is assumed to be confined to translational motion in the XH plane. The spacecraft acceleration during powered flight may be written as time invariant equations:

$$a_h = g_m (N \sin\gamma_3 - 1) \tag{1a}$$

and

$$a_x = -g_m (N \cos\gamma_3) \tag{1b}$$

where γ_3 is the flight-path angle at ignition. At ignition $t = t_3$ as defined in Fig. 1, and the respective velocity and displacement components are

$$V_h (t = t_3) = -V_3 \sin\gamma_3 \tag{2a}$$

$$V_x (t = t_3) = V_3 \cos\gamma_3 \tag{2b}$$

and

$$h (t = t_3) = h_3 \tag{3a}$$

$$x (t = t_3) \equiv 0 \tag{3b}$$

Integrating Eqs. (1a) and (1b), and substituting the ignition conditions, the following is obtained for the velocity components:

$$V_h(t_b) = -V_3 \sin\gamma_3 + g_m(N \sin\gamma_3 - 1)t_b \tag{4a}$$

and

$$V_x(t_b) = V_3 \cos\gamma_3 - g_m(N \cos\gamma_3)t_b \tag{4b}$$

where t_b is the retro motor burning time.

Similarly, by integrating Eqs. (4a) and (4b) and substituting the ignition conditions, the displacement is:

$$h(t_b) = h_3 - (V_3 \sin\gamma_3)t_b + \frac{1}{2} g_m(N \sin\gamma_3 - 1)t_b^2 \tag{5a}$$

and

$$x(t_b) = (V_3 \cos\gamma_3)t_b - \frac{1}{2} g_m(N \cos\gamma_3)t_b^2 \tag{5b}$$

The above equations describe the motion of the spacecraft along a powered descent trajectory in terms of acceleration, velocity, and displacement. Starting from a given ignition altitude, it is desirable to minimize impact velocity of the payload. The motor thrust must be terminated at some minimum altitude, however, to allow time for the descent hardware to separate from the payload. Since the retro motor is creating an acceleration about eight times that of the Martian gravity, the thrust termination point for minimum impact velocity is nearly coexistent with the point of minimum termination velocity. The minimum vehicle velocity point is, therefore, used to determine when thrust termination should occur.

Generally, the velocity may be written as

$$V^2 (t_b) = V_x^2 (t_b) + V_h^2 (t_b) \tag{6}$$

and the minimum velocity occurs when the burning time is given by:

$$t_b = \frac{V_3 (N - \sin\gamma_3)}{g_m (N^2 - 2N \sin\gamma_3 + 1)} \tag{7}$$

The burnout altitude is then given by Eq. (5a), with Eq. (7) substituted for the burning time.

Derivation of the guidance mechanization equations involves a straightforward matching of measured initial conditions to desired terminal (burnout) conditions. Such a treatment has resulted in the detailed mechanization scheme presented below. Measured and calculated range and velocity terms are identified with an asterisk to indicate possible differences from actual values. Terms identified with a bar are assumed average values of the parameters over the range of predicted trajectory variations. The subscripts i (i = 1, 2, 3, 4) used with slant range (R), velocity (V), and flight-path angle (γ) refer to the times indicated with similar subscripts in Fig. 1.

The sequence of guidance operations is as follows:

1. Measure time increment (Δt_1) between marks associated with two slant-range values R_1^* and R_2^*.

2. Calculate V_2^* as follows:

$$V_2^* = \frac{1}{\Delta t_1}\left[R_1^* - R_2^* - \overline{R\Delta\gamma \cot\gamma}\right] - \bar{\dot{V}}\frac{\Delta t_1}{2} \tag{8}$$

It should be noted that the expression $\bar{\dot{V}}$ is the average magnitude of velocity rate reduction due to drag. The minus sign has been included.

3. Calculate Δt_2 from the following quadratic:

$$\left(\bar{\dot{V}}^2 K_2 - \frac{\bar{\dot{V}}}{2}\right)\Delta t_2^2 + \left(V_2^* - \overline{R\dot{\gamma}\cot\gamma} - 2\dot{V}_2^* \bar{\dot{V}} K_2\right)\Delta t_2 + \left(R_{43} + V_2^{*2} K_2 - R_2\right) = 0 \tag{9}$$

where

$$K_2 = \frac{K_1}{g_m}\left[1 - \frac{NK_1}{2}\right] \tag{9a}$$

$$K_1 = \frac{N - \sin\bar{\gamma}_3}{N^2 - 2N \sin\bar{\gamma}_3 + 1} \tag{9b}$$

$$R_{43} = \frac{h_4 + \left(\frac{g_m t_b^2}{2}\right)_{max}}{\sin\gamma_{3_{min}}} \tag{9c}$$

T THRUST OF RETRO MOTOR
g_m MARTIAN ACCELERATION OF GRAVITY
γ FLIGHT PATH ANGLE
θ INERTIAL ANGLE
V VELOCITY

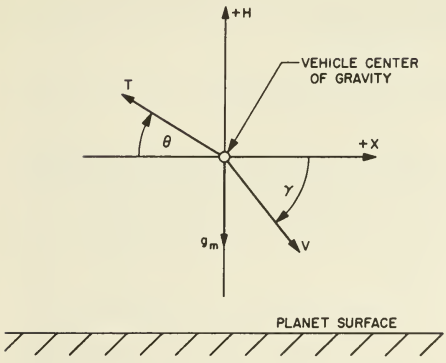


Fig. 3. Coordinate system and definitions

4. Calculate t_b as follows:

$$t_b = \left(V_2^* - \bar{V}\Delta t_2\right)\frac{K_1}{g_m} \qquad (10)$$

5. Enter sequence times by adding increments as follows:

$$t_2 = t_1 + \Delta t_1$$

$$t_3 = t_2 + \Delta t_2$$

$$t_4 = t_3 + t_b$$

The above equations represent a possible scheme for mechanizing a guidance and control system. The accuracy of these equations over the potential range of operating conditions can only be determined by exercising the equations in simulated flights. The various constant terms provide a leeway to bias the control scheme such that the errors are minimized in the range of operation which has maximum impact velocity.

Ideal System Performance

Nominal descent trajectories were selected by defining a range of allowable terminal conditions and picking a nominal point halfway between the extremes. If it is assumed that the descent propulsion system functions nominally and the burnout velocity is a minimum, then impact velocity varies primarily with burnout altitude. Minimum impact velocity is associated with the minimum allowable burnout altitude (assumed to be 400 ft for this analysis), and the maximum impact velocity is a limit which must be defined (a desirable upper limit of 200 fps was used).

Figure 4 shows burnout velocity vs altitude combinations which satisfy the specified limits on burnout altitude and impact velocity. Also plotted are ballistic trajectory altitude vs velocity curves for the range of ballistic parameters previously defined.

The allowable range of ignition conditions was determined by selecting initial conditions from the ballistic trajectories in Fig. 4 and using Eqs. (4) and (5), with burning time given by Eq. (7), to compute the minimum or burnout velocity and burnout altitude. (The terms minimum velocity and burnout velocity will be used synonymously throughout the article to define velocity at thrust termination.) The three data points on each ballistic curve of Fig. 4 represent the maximum, nominal, and minimum ignition altitudes, respectively. Values of flight-path angle (γ) are shown in parenthesis adjacent to the data points.

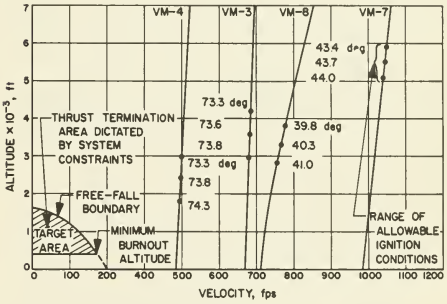


Fig. 4. Powered flight-boundary conditions

Figure 5 illustrates the correlation between ignition and burnout conditions for VM-7. The right-hand portion of the figure shows ignition velocity and altitude, while the left-hand portion shows burnout altitude and velocity and the free-fall trajectory from burnout to impact.

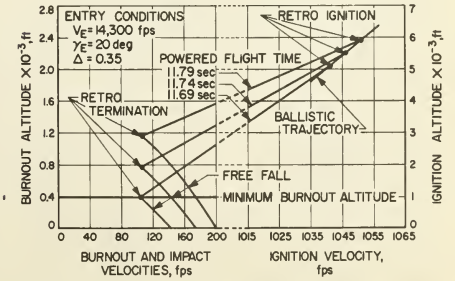


Fig. 5. Ideal terminal descent trajectories for VM-7

Table 1 presents nominal powered flight characteristics for the four atmospheric models. These data show the wide band of ignition slant range required to satisfy burnout constraints for the possible atmospheric variations. It should be noted that in no case is the ideal burnout velocity zero; it varies from 20 fps for VM-4 to 105 fps for VM-7. This result is caused by the imposition of an inertially fixed thrust mode; the downward force of gravity during the motor burn phase produces a velocity component which cannot be removed. Correction of this phenomenon would require a pitch turn during the propulsive descent phase with an obvious increase in guidance and control complexity.

The velocity increment, or ΔV required of the retro motor, may be obtained from the product of the thrust acceleration and burning time. Since the thrust acceleration is constant at approximately 97 ft/sec^2, ΔV can be obtained directly from Table 1 for each atmosphere.

Performance with Mechanization Errors

The ideal descent trajectories previously discussed indicate the range of burnout and impact conditions resulting from ballistic trajectory variations. Additional variations resulting from sensing and execution errors are discussed in this section. The sources of error are confined to those occurring from propulsion and radar system uncertainties. Errors such as thrust vector misalignment errors and spacecraft pointing errors are not considered at this time. These errors are small if an active attitude control system is used during the sensing and ignition period (t_1 through t_3) to damp out aerodynamic oscillations. Effects of steady winds are discussed at the end of the paper.

The guidance and control analytical model presented previously was not used in its entirety for the error analysis because of the complex form of the equations. Certain assumptions were made to allow a preliminary evaluation of gross system performance without entering the complexities of correlated error coefficients. The following assumptions were made:

1. The velocity, slant range, and path angle required for ignition are known prior to ignition.

2. Only two slant-range measurements are necessary to determine the ignition conditions, excluding path angle which is determined independently of the radar.

3. At the second slant-range measurement the ignition requirements are met, and ignition occurs after a negligible elapsed time.

Assumptions 1 and 2 are consistent with the analytical model previously derived except that a constant value of path angle was assumed in the model. If a constant value is assumed near the actual value associated with the worst-case trajectory, the results should be a flattening of the maximum impact velocity variation with atmosphere.

The third assumption tends to modify the correlation between velocity and range prediction errors. Coincidence of the second range mark and ignition point makes the range errors identical for the two points. For the model previously developed, however, a late second range mark, coupled with an exact mark at R_1, would magnify the ignition error. (A low estimate of velocity would make Δt_2 larger than nominal.) The assumption was made to simplify the analysis with the recognition that the results so obtained must be appropriately qualified.

For the original analytical model:

$$V_3^* = \frac{1}{\Delta t_1}\left[R_1^* - (R_2^* + R\Delta\gamma\cot\gamma)\right] - \bar{V}\left(\frac{\Delta t_1}{2} + \Delta t_2\right) \qquad (11)$$

Table 1. Nominal powered flight conditions

Atmosphere	VM-3	VM-4	VM-7	VM-8
Ignition Conditions				
Altitude, ft	3600	2400	5500	3300
Flight path angle, deg	73.6	73.8	43.7	40.3
Slant range, ft	3750	2500	7970	5080
Velocity, fps	680	496	1046	768
Burning time, sec	8.00	5.83	11.74	8.56
Burnout conditions				
Altitude, ft	950	1000	780	880
Velocity, fps	28	20	105	82
Impact velocity, fps	156	158	173	168

The assumptions just discussed create the following equation which was used for the error analysis presented herein.

$$V_3^* = V_2^* = \frac{1}{\Delta t_1}\left[R_1^* - R_2^{c*}\right] \qquad (12)$$

where

$$R_2^{c*} = R_2^* + R_1^* \Delta\gamma \cot\gamma + \dot{\overline{V}} \frac{\Delta t_1^2}{2} \qquad (13)$$

Radar-Induced Errors. Ignition range dispersions result directly from errors in the radar's range measurement; these errors translate into burnout dispersions and are compounded by the motor burn time, which is based on the range marks. The Radar system error sources were obtained from the Ranger Block V Altimeter System Analysis (4). These error sources were chosen as being representative of errors expected in the assumed radar system. Specific values in the reference are based on a parametric operating range of 70,000 ft, a velocity of 9,000 fps, and flight-path angles up to 30° off vertical. Transformation of these errors to the operating range of interest and further identification of the individual error sources for the narrow beam antenna case follow:

1. Electrical Error — The electrical system errors are generally fuzing or response time errors and are independent of path angle but proportional to velocity. In Reference 4, the 1-σ electrical error is 48 ft. If this is rewritten as a function of velocity

$$\sigma_E = \frac{48}{9000} V = 5.33 \times 10^{-3} V \text{ feet} \qquad (14)$$

2. Ranging Error — The ranging error is the error in the basic timing of the range pulse; i.e., elapsed time error of the received signal. In Reference 4, a sensitivity of 490 ft/μsec or one-half the speed of light, was specified. If a 0.09-μsec, 1-σ error in rise time is assumed, the ranging error is

$$\sigma_{RG} = 44.1 \text{ feet} \qquad (15)$$

This error is inherent in the radar system and, consequently, independent of the measurable parameters.

3. Boresight Alignment Error — The boresight alignment error is a small angular error resulting when the antenna boresight direction and the flight-path direction do not coincide. If $\Delta\gamma_B$ is the boresight error, ΔR may be expressed as

$$\Delta R = h\left[\frac{1}{\sin\gamma} - \frac{1}{\sin(\gamma + \Delta\gamma_B)}\right]$$

Expanding $\sin(\gamma + \Delta\gamma_B)$ and simplifying, the following is obtained:

$$\Delta R = R\left(\frac{\Delta\gamma_B}{\Delta\gamma_B + \tan\gamma}\right)$$

If a 1-σ boresight error of 0.22 deg or 3.84 mr is assumed, the 1-σ error in range becomes

$$\sigma_{BS} = R\left(\frac{1}{1 + 260\tan\gamma}\right) \text{ feet} \qquad (16)$$

4. Antenna Beamwidth Error — This error is essentially the ambiguity associated with defining when fuzing occurs relative to the time between the first appearance of a lower beam edge return in the range gate and the moment the peak signal appears in the range gate. Two components of error result from the ambiguity: a bias component and a random component. The bias component will be deleted from consideration, although it may be extremely large, because it can be accounted for in the gate setting.

If it is assumed that fuzing will occur within 0.5 deg, 1 σ of the equivalent boresight angle of the lower beam edge, the 1-σ error in range due to beam width error is

$$\sigma_{BW} = R\left[\frac{0.0087 \cos\gamma}{\sin\gamma + \theta_{1/2}\cos\gamma}\right] \qquad (17)$$

where $\theta_{1/2}$ = one half of the antenna beam width. For calculation purposes, a 2-deg half-width antenna will be assumed.

If it is assumed that the error sources are uncorrelated and normally distributed, a total 1-σ radar system error may be computed by taking the root-sum-square of the individual sources. Therefore, the 1-σ error in measuring slant range is

$$\sigma_R = \left(\sigma_E^2 + \sigma_{RG}^2 + \sigma_{BS}^2 + \sigma_{BW}^2\right)^{1/2} \qquad (18)$$

After the radar errors are known, ignition dispersions can be calculated. If the measured value notation (*) is deleted and the subscript from Δt is dropped, the ignition velocity is approximately given by

$$V_3 = \frac{1}{\Delta t}\left(R_1 - R_2^c\right) \qquad (19)$$

Expansion in a first-order Taylor series yields

$$dV_3 = \frac{\partial V_3}{\partial R_1} dR_1 + \frac{\partial V_3}{\partial R_2^c} dR_2^c + \frac{\partial V_3}{\partial \Delta t} d\Delta t \qquad (20)$$

If it is assumed that dR_1 and dR_2^c are independent, normally distributed, random variables, and $d\Delta t$ is negligible, the variance of the ignition velocity is

$$\sigma_{V_3}^2 = \frac{1}{\Delta t^2}\left[\sigma_{R_1}^2 + \sigma_{R_2^c}^2\right] \quad (21)$$

where σ_{R_1} and $\sigma_{R_2^c}$ are the 1-σ errors obtained in measuring slant range. Similarly, the variance of the ignition altitude is

$$\sigma_{h_3}^2 = \left(\sin^2\gamma_3\right)\sigma_{R_3}^2 + \left(R_3^2\cos^2\gamma_3\right)\sigma_{\gamma_3}^2 \quad (22)$$

where σ_{γ_3} is the 1-σ error in determining the flight-path angle.

Equations (21) and (22) were evaluated for the four atmospheres at the nominal ignition condition using the radar errors derived and a 3-σ value of 1 deg for the path-angle error. The resulting 1-σ dispersions in ignition conditions are shown in Table 2 for Δt intervals of 0, 5, 1, and 5 sec. The velocity dispersions are reduced considerably as Δt increases, but the 1-σ error in ignition altitude is independent of Δt and, hence, remains fixed. Analytically, as the time interval increases, the ignition dispersions become confined to those dispersions in ignition altitude only.

Dispersions in burnout altitude and velocity will be caused by measurement and approximation errors associated with the determination of the ignition conditions. Whether or not these dispersions will produce burnouts or landings outside the constraint values is of primary importance.

Table 2. One-sigma velocity measurement and ignition altitude errors

Error		Atmosphere			
		VM-3	VM-4	VM-7	VM-8
σ_h (ft)		45	44	75	52
σ_V (fps)	Δt_1 = 0.5 sec	130	128	256	206
	Δt_1 = 1 sec	65	64	132	106
	Δt_1 = 5 sec	13	13	32	26

The dispersions of the horizontal and vertical components of the burnout velocity may be expressed as linear functions of the dispersions of V_3 and γ_3 using partial derivatives of Eqs. (4a) and (4b).

$$\begin{bmatrix} dV_x \\ dV_h \end{bmatrix} = \begin{bmatrix} \cos\gamma_3 & (V_3 - g_m N t_b)\sin\gamma_3 \\ -\sin\gamma_3 & (g_m N t_b - V_3)\cos\gamma_3 \end{bmatrix} \begin{bmatrix} dV_3 \\ d\gamma_3 \end{bmatrix} \quad (23)$$

If it is assumed that dV_3 and $d\gamma_3$ are independent, normally distributed, random variables with zero means, the following can be written for the covariance matrix of the velocity components:

$$\Lambda = \begin{bmatrix} \sigma_{V_x}^2 & \rho\sigma_{V_x}\sigma_{V_h} \\ \rho\sigma_{V_x}\sigma_{V_h} & \sigma_{V_h}^2 \end{bmatrix} = P\Omega P^T \quad (24)$$

where P is the matrix of partial derivatives defined in Eq. (23) Ω is the covariance matrix of the random variables:

$$\Omega = \begin{bmatrix} \sigma_{V_3}^2 & 0 \\ 0 & \sigma_{\gamma_3}^2 \end{bmatrix} \quad (25)$$

where P^T is the transpose of P, and ρ is the correlation coefficient of V_x and V_h.

Substituting for the partial derivatives in the P matrix and performing the indicated operations in Eq. (24), the covariance matrix becomes

$$\begin{bmatrix} \sigma_{V_x}^2 & \rho\sigma_{V_x}\sigma_{V_h} \\ \rho\sigma_{V_x}\sigma_{V_h} & \sigma_{V_h}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (26)$$

where the matrix elements are defined as

$$a_{11} = \cos^2\gamma_3\,\sigma_{V_3}^2 + \sin^2\gamma_3 \times \left(V_3 - g_m N t_b\right)\sigma_{\gamma_3}^2$$

$$a_{12} = a_{21} = \frac{\sin 2\gamma_3}{2}\left[\left(V_3 - g_m N t_b\right)^2 \times \sigma_{\gamma_3}^2 - \sigma_{V_3}^2\right]$$

$$a_{22} = \sin^2\gamma_3\,\sigma_{V_3}^2 + \cos^2\gamma_3 \times \left(V_3 - g_m N t_b\right)^2\sigma_{\gamma_3}^2$$

Thus the variance of the velocity components is defined by a_{11} and a_{22} and the correlation coefficient of V_x and V_h is given by

$$\rho = \frac{\sin 2\gamma_3}{2\sigma_{V_x}\sigma_{V_h}}\left[\left(V_3 - g_m N t_b\right)^2\sigma_{\gamma_3}^2 - \sigma_{V_3}^2\right] \quad (27)$$

The term $\left(V_3 - g_m N t_b\right)^2 \sigma_{\gamma_3}^2$ is negligible for the assumed value of σ_{γ_3}, and, hence, the correlation coefficient may be written as

$$\rho \approx - \frac{\sigma_{V_3}^2 \sin 2\gamma_3}{2\sigma_{V_x}\sigma_{V_h}} \tag{28}$$

The dispersion in burnout altitude can be expressed as a function of h_3, V_3, and γ_3 dispersions using the partial derivatives of Eq. (5a)

$$\begin{bmatrix} dh \end{bmatrix} = \begin{bmatrix} 1 & -t_b \sin\gamma_3 & t_b \cos\gamma_3 \left(\frac{1}{2} g_m N t_b - V_3\right) \end{bmatrix} \begin{bmatrix} dh_3 \\ dV_3 \\ d\gamma_3 \end{bmatrix} \tag{29}$$

Again, if it is assumed that dh_3, dV_3 and $d\gamma_3$ are independent, normally distributed, random variables, the variance of the burnout altitude is

$$\sigma_h^2 = \sigma_{h_3}^2 + t_b^2 \sin^2\gamma_3 \sigma_{V_3}^2 + t_b^2 \cos^2\gamma_3 \times \left(1/2\, g_m N t_b - V_3\right)^2 \sigma_{\gamma_3}^2 \tag{30}$$

One-sigma errors in burnout velocity and altitude resulting from errors in the ignition conditions are defined by Eqs. (26) and (30). These errors were evaluated for the nominal ignition conditions, and 1-σ errors in velocity and altitude are plotted as a function of sampling interval Δt in Fig. 6. The resulting distribution function of each of the random variables V_x, V_h, and h is Gaussian and, therefore, the range of expected variation may be easily determined if the mean and standard deviation of the distribution is known.

To evaluate the effect of the radar system errors on system feasibility, the impact velocity distribution function is also desired. Impact velocity is related to burnout conditions by

$$V_I^2 = V_{bo}^2 + 2 g_m h_{bo} \tag{31}$$

where $V_{bo}^2 = V_x^2 + V_h^2$. The distribution functions of the random variables in Eq. (31) and the correlation of V_x and V_h are known. However, this information alone is insufficient to determine the burnout velocity distribution function. Other factors, such as the correlation between burnout altitude and velocity, must be analyzed before a distribution function satisfying the non-linear impact velocity equation can be obtained. Without pursuing this line of analysis further, a Monte Carlo technique will be used for determining the impact velocity distribution. The analysis and results are presented in the section on Monte Carlo Analysis of Impact Velocity.

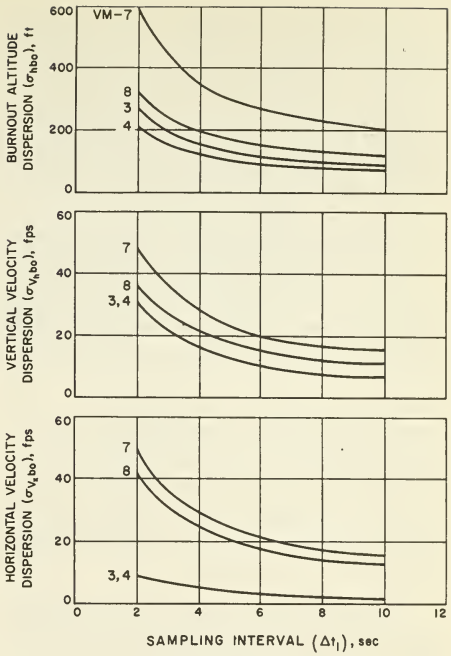


Fig. 6. One-sigma dispersions in burnout altitude and burnout velocity components as a function of sampling interval

For comparison with the Monte Carlo results, Gaussian distribution functions of burnout altitude are shown in Fig. 7 for the worst-case atmosphere, VM-7. Distribution functions are plotted for three sampling times; a maximum range of expected burnout altitudes can be obtained for a given probability.

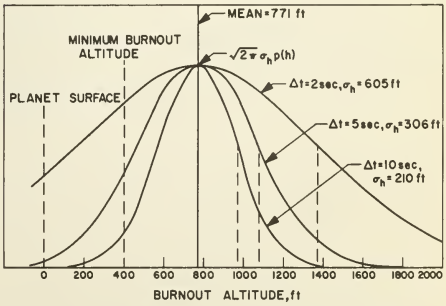


Fig. 7. Burnout altitude distributions, VM-7

Propulsion-Induced Errors. The propulsion system assumed for this lander model is a fixed-thrust, solid-propellant motor with variable total impulse. Propulsion system errors, which result in burnout velocity and altitude uncertainties, were considered for two types of impulse control systems: an accelerometer system and the timer system. Both systems were analyzed in sufficient detail to determine how propulsion system variations, under certain given assumptions, influence system performance.

The accelerometer system integrates the acceleration to obtain velocity and causes thrust termination when the velocity increment reaches a value determined by the guidance computer. Included is a ΔV bias which accounts for the tail-off characteristics of the retro motor and elapsed sequencing time. Sources of error include accelerometer errors, resolution and scale-factor errors, and variations in thrust-termination impulse.

The timed system produces a thrust-termination signal at a computed interval after motor ignition. Motor tail-off characteristics and sequencing time are accounted for in the system. The primary source of error for this system is an uncertainty in thrust level caused by temperature variations and normal motor-to-motor variations. The errors in timer accuracy and other mechanization errors were considered negligible.

Dispersions in burnout velocity and displacement resulting from propulsion system uncertainties were evaluated for the normal ignition conditions for both the accelerometer and timed thrust-termination systems. Table 3 lists the $1-\sigma$ error values assumed, and the resulting burnout dispersions are tabulated in Table 4. Error equations similar to those derived for the radar errors were used to calculate the values shown. As in the radar analysis, the largest errors occurred for atmosphere VM-7. Comparison of the propulsion errors to the radar errors, however, shows an order-of-magnitude difference. Propulsion errors were therefore not considered further.

Table 3. One-sigma error values assumed

Accelerometer System

$\sigma_{\Delta V}$ = 0.2% caused by electronic-induced errors and shutdown impulse uncertainties

Timed System

$\sigma_{\Delta V_1}$ = 0.5% caused by temperature uncertainties

$\sigma_{\Delta V_2}$ = 0.2% caused by impulse variation at a known temperature

Monte Carlo Analysis of Impact Velocity. On the basis of the analysis presented in the section on Radar-Induced Errors, a Monte Carlo technique was employed to determine the range of expected impact velocities. Without this information, system feasibility may only be partially analyzed in terms of burnout altitude and velocity components dispersions.

Propulsion errors were shown to be relatively insignificant and, therefore, were not considered in the Monte Carlo study. It is believed that feasibility is adequately evaluated with these errors deleted.

The Monte Carlo analysis was performed for the nominal ignition cases using Eqs. (23) and (29). Random samples of h_3, V_3, and γ_3 were selected from Gaussian populations scaled by the proper mean value and standard deviation. The impact velocity was computed from the relation

$$V_I^2 = V_x^2 + V_h^2 + 2g_m h \qquad (32)$$

For each case evaluated, the burnout altitude and impact velocity were obtained for 100 random samples.

Table 4. One-sigma dispersions in burnout altitude and velocity components

Atmosphere	Accelerometer System			Timed System		
	V_x, fps	V_h, fps	h, ft	V_x, fps	V_x, fps	h, ft
VM-3	0.44	1.57	17.2	1.16	3.95	15.8
VM-4	0.31	1.15	9.2	0.84	2.89	8.4
VM-7	1.64	1.75	31.1	4.37	4.18	24.5
VM-8	1.26	1.21	15.9	3.36	2.85	12.2

Distribution functions for impact velocity and burnout altitude are given in Figs. 8 and 9 for the two bounding atmospheric models (VM-4 and VM-7). Cumulative distribution functions of impact velocity and burnout altitude are shown in Figs. 10 and 11 for VM-7.

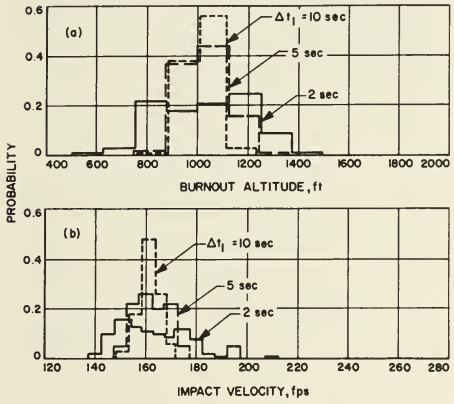


Fig. 8. Burnout altitude and impact velocity distribution functions, VM-4

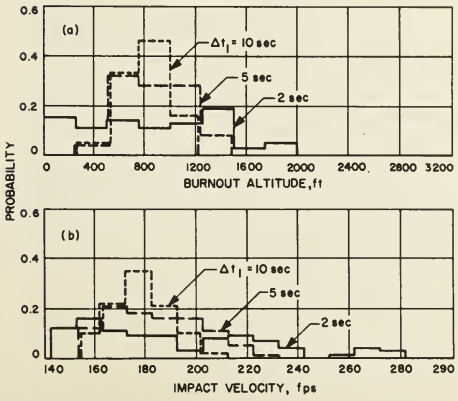


Fig. 9. Burnout altitude and impact velocity distribution functions, VM-7

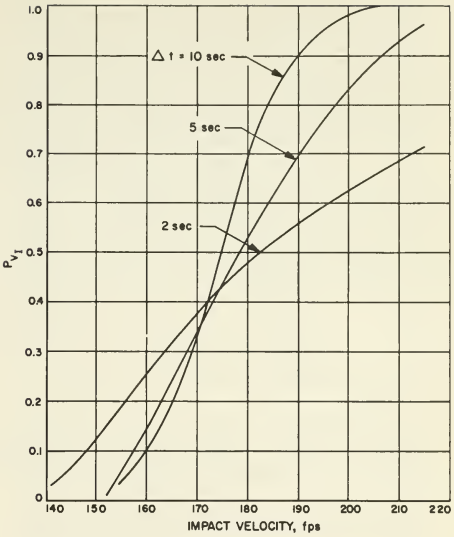


Fig. 10. Approximate cumulative distribution functions of impact velocity, VM-7

Evaluation of the Monte Carlo results, in conjunction with the radar error distributions previously discussed leads to the following conclusions:

1. The range measurement time interval (Δt_1) should be at least 5 sec long so as to keep measurement errors from producing unacceptably large impact velocities.

2. Significant variations exist in impact velocity distributions among the atmosphere models considered; the largest errors are associated with the largest initial velocities. These variations should allow assumption of an average flight-path angle for the mechanization equations without significantly affecting the maximum impact velocities. Use of fixed range marks in an actual mechanization makes Δt inversely proportional to velocity; accuracy is thereby improved for the slower velocity cases, and the allowance for parameter bias is increased.

3. Within the context of the assumptions discussed in the section on Performance with Mechanization Errors, the analyses indicate that the system is close to performing within the boundary conditions established. For example, take 96% probability as an acceptable level, and let Δt = 5 sec, the VM-7 case gives a maximum impact velocity of 215 fps and altitude of 400 ft (Figs. 10 and 11). If a higher probability level is desired, however, the nominal point must be raised to get the minimum burnout altitude above 400 ft, or a longer Δt must be selected to reduce errors.

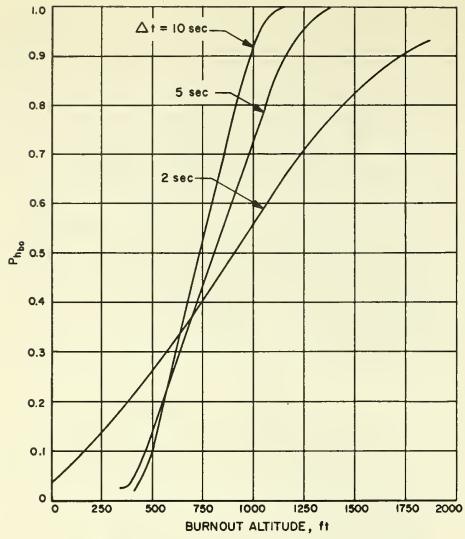


Fig. 11. Approximate cumulative distribution functions of burnout altitude, VM-7

System Performance Summary

Vehicle pointing errors were not included in the analysis of the previous section on Performance with Mechanization Errors because it was assumed that the vehicle attitude would be controlled to the steady-state aerodynamic direction (zero-angle-of-attack). Under such control, dispersions caused by pointing errors about the nominal are small compared to potential steady-state wind effects. If it is assumed that the local atmosphere is moving with respect to the ground at a steady wind velocity V_ω from the time of the first range mark to burnout (t_1 to t_4), the same effect is achieved if the ground is moving with a velocity $-V_\omega$ with respect to the atmosphere. Since the velocities calculated in this paper are all based on atmospheric references (the body is aligned to local aerodynamic velocity), the steady-state wind velocity must be added to the impact velocities previously calculated.

Vector addition of the two velocity components will give quite a range of values because of (1) the range of impact flight-path angles implied in the error analysis previously discussed, and (2) the probable direction of the wind. The application of a probability distribution to these factors is beyond the scope of this paper.

However, a possible solution, which tends to locate the range of values, is readily attainable. If a VM-7 trajectory is assumed, in which the errors combine to cause a shorter burning time (low ΔV), the impact flight-path angle will tend toward vertical. Assuming a vertical impact velocity of 215 fps, added to a horizontal wind of 220 fps, gives a resultant impact velocity of 308 fps. The range of values for nonvertical impact would span this value as a function of wind direction.

Therefore, the worst-case impact velocity, upon which the impact limiter must be designed, will be somewhat greater than the 250- to 300-fps values associated with a parachute system.[2] Relatively poor performance, as compared to the performance of a parachute, suggests that possible ways of improving the system should be investigated. Figure 5 shows that the minimum impact velocity, with no mechanization errors, is 140 fps in the VM-7 atmosphere. This value, compared to 120 fps for the parachute (2), shows that the basic system mechanization must be changed. This could be accomplished by using a pitch turn during retro burn; obvious complications which would immediately arise are: (1) how is "down" determined? (2) what turn rate should be used to span the range of atmosphere models?

Another way of decreasing impact velocity in the inertial hold mode would be to increase thrust levels. Such a scheme would require multiple rocket motors to achieve a significant thrust increase. The preferred mode of ΔV control would probably be to drop the payload out from between the motors at t_4 rather than to terminate each motor. This design has the added advantage that the still burning propulsion system provides positive separation between the payload and descent system.

Even if the system were improved to decrease the atmosphere-referenced impact velocity to from 50 to 100 fps, the wind velocity must be accounted for. Therefore, a system which operated perfectly based on an aerodynamic reference, and arrived at the surface with zero velocity with respect to the atmosphere, could be moving laterally at a velocity of 220 fps with respect to the ground. The wind velocity is, therefore, a lower limit on impact velocity unless a radar system, which senses velocity components with respect to the ground is used. Such a system is described in the second half of this paper.

III. Part II: Soft-Landing System

Terminal Descent System Mechanism Concept

The analysis of the rough-landing system illustrated the difficulty of landing sizable payloads on Mars using a conceptually simple terminal descent scheme. The uncertainties in the landing conditions with such a system justify the need for investigating terminal descent systems of a more complex nature which can soft-land payloads. With the successful demonstration of a lunar soft landing, the Surveyor terminal descent system concept is valuable as a starting point in the descent system design for a Martian soft-landing capsule. However, because Mars has an atmosphere with relatively unknown characteristics, the terminal descent phase conditions cannot be predicted as in the case of Surveyor and, consequently, increased complexity is necessary if the required operational flexibility is to be obtained.

In comparison with the relatively passive rough-landing terminal descent system, the requirements for a soft landing dictate an active terminal descent system whereby the position and velocity of the capsule is determined throughout the retro-controlled descent. Active control is maintained by utilizing the position and velocity information to provide the attitude control and thrust force modulation necessary to effect the landing. For the capsule considered, the basic on-board descent system was assumed to include:

1. A guidance system which determines the thrust program for soft landing.

2. A liquid propulsion system with three differentially throttlable engines, spaced 120 deg apart in a radial array.

3. A radar system for measuring doppler velocity and slant range.

4. An attitude control system for capsule stabilization and thrust vector position control.

Other systems, such as communication, power, etc., must be included to complete the entire composition of the descent system.

Upon de-orbiting from the spacecraft and entering the planetary atmosphere, the flight profile of the capsule may be discussed in terms of two distinct phases: (1) a ballistic descent phase, and (2) a retro-controlled descent phase. During the ballistic descent phase, the aeroshell provides sufficient aerodynamic braking such that the velocity is subsonic or near subsonic at the initiation of the retro-controlled flight phase. The ability to soft-land payloads of significant size is directly influenced by the effectiveness of the aerodynamic braking mechanism; the higher the ballistic velocity becomes at retro ignition, the more the propulsion system weight detracts from the payload capability of the capsule. For this reason, much consideration must be given to the design of the aeroshell.

Several control modes are necessary in the retro-controlled flight phase. The first of these, a stabilization and acquisition mode, is necessary to damp out any ballistic oscillations and ignition transients and to allow sufficient time for the on/board radar system to acquire and lock-on the planet surface and become operationally reliable. A mid-range thrust force is commanded during this mode to provide adequate torquing capability. Once the radar is reliable, the capsule begins its active descent to the surface in a gravity-turn steering mode. Gravity-turn steering implies that the thrust vector is oppositely aligned to the velocity vector, as determined by the radar system, and that the capsule will be in an upright orientation upon landing. The thrust required for this mode is determined from the guidance system.

Near the surface, the engine plume may create surface disturbances which affect the landing system. To circumvent such occurrences, a terminal mode which shuts the engines off at a specified condition above the surface may be required. The terminal mode is not considered as a descent mode in this paper, although further mention of its utility is made.

Low-Altitude Ballistic Trajectories

The ballistic trajectory defines that portion of the capsule flight from atmospheric entry to retro ignition. Atmospheric entry is assumed to occur at 800,000 ft above the Martian surface and the conditions at entry (entry velocity, entry angle, and entry angle-of-attack) define the initial conditions of the ballistic descent. These conditions, along with the particular atmospheric conditions selected, provide the inputs for determining the ballistic trajectory.

The particular entry "corridor" established for this study included an entry velocity of 15,000 fps, an entry angle of 14 to 20 deg, and an entry angle-of-attack of ± 5 deg. The entry angle bounds are determined by the "skip limit", the entry angle below which the capsule attains positive flight-path angles, and the ballistic velocity magnitude which is, in the retro ignition altitude region, a function of entry angle magnitude. The entry angle-of-attack range is a measure of the capsule's ability to align its roll-axis with the velocity vector at entry.

It is desirable for landing to impose rigid constraints on the entry conditions, particularly on the entry angle and entry angle-of-attack. Entry angle variations affect retro ignition conditions. Large deviations from a zero angle-of-attack at entry may result in prominent aerodynamic oscillations during the ballistic descent which, upon retro ignition, exceed the damping capability of the attitude control system. This effect dictates a capsule configuration which, in general, provides excellent aerodynamic properties.

The geometrical shape of the capsule considered in this study was a 60-deg blunted-cone fore-body and a frustum of a 30-deg cone aft-body. For the range of entry conditions chosen, this particular shape lends itself well to the soft-landing application. Consequently, negligible lifting forces result and capsule atmospheric alignment is maintained throughout the ballistic descent phase. Other capsule parameters of interest include a diameter of 16.1 ft, a ballistic coefficient (M/C_dA) of 0.3 slugs-ft^2, and a mass of 84.5 slugs.

Atmospheric variations were assumed bounded by Martian models VM-4, VM-7, VM-8, and VM-10.

The first three were the models used in the rough-landing system study and VM-10, a 20-mb atmosphere, was included to demonstrate a soft-landing capability under a worst-case expected atmosphere.

Using the above parameters as input data to a 3-degree-of-freedom capsule ballistic descent computer simulation, ballistic trajectories from 800,000 ft to the surface were computed. A 6-degree-of-freedom simulation had previously shown that the capsule configuration chosen for this study was aerodynamically stable and its motion was essentially 3-degree-of-freedom motion.

The low-altitude ballistic descent trajectories are shown in Fig. 12. An altitude-aerodynamic velocity characteristic is plotted for each of the atmospheric models and entry angles considered. Aerodynamic flight-path angles, referenced to the horizontal, are shown adjacent to the circled data points. The trajectories are illustrated for a maximum altitude of 25,000 ft. Previous studies have shown that confinement to this range is sufficient to determine the retro ignition conditions.

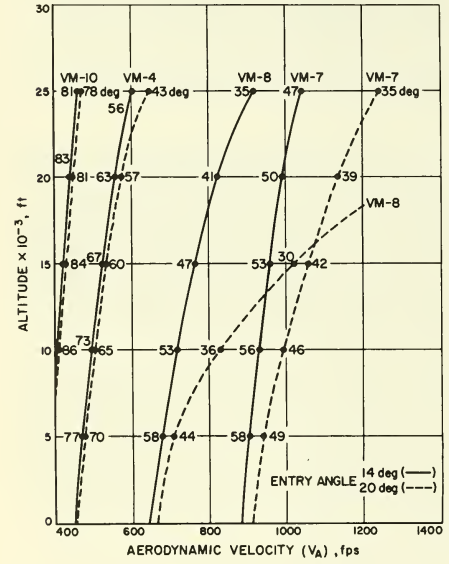


Fig. 12. Low-altitude ballistic trajectories

Steady-State Wind Effects

Prior to ignition, the capsule remains aligned with the aerodynamic velocity vector, V_A, defined as the velocity of the capsule with respect to the planetary atmosphere. This alignment capability is provided by the aerodynamic forces acting on the capsule. In the case of no steady-state wind, the surface-referenced velocity vector, V_g, the velocity of the capsule with respect to the surface, is aligned with V_A. In the case of a steady-state wind velocity parallel to the surface of the planet, V_g will be a function of the steady-state wind velocity, while V_A will remain unaffected since the capsule flies with the wind prior to ignition.

Upon ignition, the stabilization and acquisition mode is initiated. During this mode, the attitude control system maintains the capsule's pre-ignition alignment. The gravity-turn mode is then initiated and the attitude of the capsule is altered such that the thrust vector is oppositely aligned with V_g. This alignment is maintained until completion of the gravity-turn mode.

Retro Ignition Conditions

Retro ignition conditions may be defined as the altitude, aerodynamic velocity, and the aerodynamic path angle at which retro engines are ignited. Ideally, ignition should occur after most of the entry velocity has been ballistically removed. Igniting at excessive altitudes may present fuel consumption problems and, conversely, igniting too close to the surface may result in thrust requirements which exceed the propulsion system capabilities. Although precise control of the ignition altitude is not necessary, ignition altitude should be prescribed within a range compatible with the descent system limitations and the atmospheric uncertainties.

In this study, the criterion established for retro ignition was to ignite the engines when the slant range equalled a pre-selected value. The implementation of this criterion requires a range-marking radar (RMR) or some other device which generates a marking signal when the particular slant-range value is reached. The marking signal is then sent to the retro engines, and ignition occurs. A slant range of 20,000 ft was chosen as the ignition slant range. Retro ignition conditions are tabulated in Table 5 for the two entry angles and the atmospheric models considered. The data represent the ignition conditions for the case of a steady-state wind, parallel to the surface, and in the direction of horizontal motion. Table 5 shows that significant variations exist in the ignition conditions over the atmospheric model range, particularly in the ignition altitude. This is predominantly a result of the sensitivity of the ignition criterion to atmospheric density.

Terminal Descent Guidance System Analysis

Equations of Motion. The equations of motion describing the capsule descent may be derived by considering the force diagram shown in Fig. 13. It is assumed here that the capsule is descending in the gravity-turn mode, under retro power, and that the motion is essentially specified by considering the motion in the vertical descent plane. The four forces shown acting on the capsule are the two aerodynamic forces, axial, F_A, and normal, F_N, the capsule weight, and the engine thrust force, F_T. Resolution of these forces into their coordinate system components will yield the equations of motion.

The resultant forces acting in the +X and +h direction, respectively, are given by

$$F_x = F_N \sin \gamma_g - (F_A + F_T) \cos \gamma_g \qquad (33)$$

Table 5. Retro ignition conditions

Atmosphere	Altitude, h (ft)	Aerodynamic velocity, V_A (ft/sec)	Surface-referenced velocity, V_g (ft/sec)	Aerodynamic flight-path angle, γ_A (deg)	Surface-referenced flight-path angle, γ_g (deg)	Steady-state wind velocity, V_W (ft/sec)
Entry Angle = 14 deg, Entry Velocity = 15,000 ft/sec						
VM-4	18,200	535	582	65.4	53.7	155.5
VM-7	15,700	965	1020	52	44.8	220
VM-8	14,600	770	830	47	39.3	220
VM-10	19,900	435	440	83.1	69.0	110
Entry Angle = 20 deg, Entry Velocity = 15,000 ft/sec						
VM-4	17,000	540	590	58.7	47.5	155.5
VM-7	13,800	1050	1095	43.8	38.1	220
VM-8	11,400	865	935	34.9	30.2	220
VM-10	19,700	440	445	81.1	67.5	110

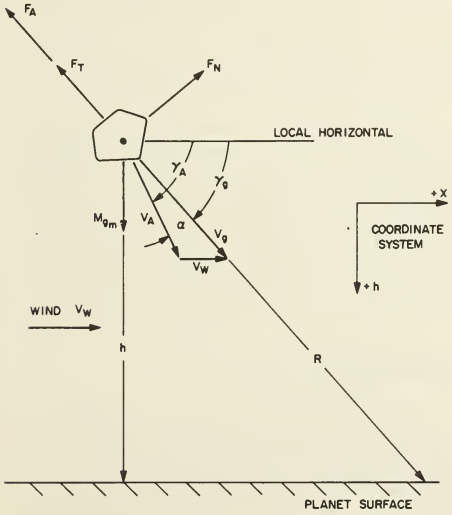


Fig. 13. Capsule force-velocity diagram in the vertical descent plane

and

$$F_h = -F_N \cos \gamma_g - (F_A + F_T) \sin \gamma_g + M g_m \quad , \tag{34}$$

where γ_g is the surface-referenced flight-path angle, M is the capsule mass, and g_m is the Martian acceleration of gravity. (For descent during the stabilization and acquisition mode, γ_g should be replaced by γ_A.) Normalizing the equations by the capsule mass, the differential equations for the velocity components may be expressed as

$$\frac{dV_x}{dt} = a_x = a_N \sin \gamma_g - (a_A + A_T) \cos \gamma_g \tag{35}$$

and

$$\frac{dV_h}{dt} = a_h = -a_N \cos \gamma_g - (a_A + A_T) \sin \gamma_g + g_m \quad . \tag{36}$$

The final differential equation necessary to determine position and velocity of the capsule as a function of time is

$$\frac{dh}{dt} = -V_g \sin \gamma_g \quad . \tag{37}$$

Each of the above equations are implicit functions of time and are only integrable by numerical methods.

The altitude is determined by numerically integrating Eqs. (35), (36), and (37), and the surface-referenced velocity is computed from the equation

$$V_g = \left(V_x^2 + V_h\right)^{1/2} \tag{38}$$

and the surface-referenced path angle is given by

$$\gamma_g = \tan^{-1}\left(\frac{V_h}{V_x}\right) \tag{39}$$

In the case of a steady-state wind, as shown in Fig. 13, the angle-of-attack may be expressed as

$$\alpha = \tan^{-1}\left(\frac{V_h}{V_x - V_w}\right) - \gamma_g \tag{40}$$

In Eqs. (35) and (36), a_N and a_A represent, respectively, the normal and axial aerodynamic accelerations and are given by

$$a_N = \frac{\rho V_A^2 S C_N}{2M} \tag{41}$$

and

$$a_A = a_N \frac{C_A}{C_N} \tag{42}$$

where C_N and C_A are the normal and axial force coefficients, ρ is the atmospheric density (assumed to vary exponentially), and S is the capsule reference area. The values of C_N and C_A are determined from the aerodynamic data of the capsule which describe these coefficients as functions of Mach number and angle-of-attack.

The thrust acceleration of the retro engines is denoted by A_T in Eqs. (35) and (36).

Terminal Descent Guidance Law. The basic descent guidance concept involves gravity-turn steering throughout the retro-controlled descent, with the exception of the stabilization and acquisition mode immediately after ignition. Gravity-turn steering is desirable because, as the velocity of the capsule approaches zero, the capsule orientation tends toward the vertical. Consequently, no violent re-orientation maneuver is required prior to landing.

Initial efforts to derive a gravity-turn steering guidance law for thrust acceleration control yielding a terminal condition of zero velocity at zero altitude resulted in the expression

$$A_T' = \frac{V_I^2}{4h_I}\left\{\sin\gamma_I + \left[\sin^2\gamma_I + \frac{4g_m h_I}{V_I^2}\left(\frac{4g_m h_I}{V_I^2} + \sin^2\gamma_I + 1\right)\right]^{1/2}\right\} \tag{43}$$

where the subscript I denotes the conditions at ignition. This expression was derived under the assumption that (1) no atmosphere was present, (2) the thrust acceleration remained constant throughout the retro-controlled descent, and (3) Martian gravity, g_m, was a constant. The validity of the law was limited to a flight-path angle range of $0 \text{ deg} \le \gamma_I \le 90$ deg.

Because of the mathematical complexity of this law, a complex mechanization is required for implementation. Therefore, a simplification was sought which resulted in the same zero-zero landing conditions. This simplification was valid in the presence of an atmosphere, but did not necessarily restrict the thrust acceleration level to a constant. Considering the vertical descent case ($\gamma_I = 90$ deg), which is a limiting form of Eq. (43), A_T' becomes

$$A_T' = g_m + \frac{V_I^2}{2h_I} \tag{44}$$

If the subscript I is deleted and the altitude, h, is replaced by the slant range, R, then Eq. (44) becomes

$$A_T = g_m + \frac{V^2}{2R} \tag{45}$$

where A_T is the instantaneous thrust acceleration required to soft-land, and V and R are the instantaneous velocity and slant range. As a result of the higher differential change in R than in h, the thrust acceleration increases during the descent and asymptotically approaches the value of A_T in Eq. (43) as the orientation approaches vertical. (In the nomenclature of the previous section, V is the surface-referenced velocity, V_g.)

From previous soft-landing studies which utilized the expression for A_T as the basic guidance law, all results indicated that suitable landing conditions can be achieved for the majority of environmental conditions on Mars. However, for the more dense atmospheric models, some difficulty is encountered because excessive fuel is required to land the capsule in these atmospheres. The proposed mechanization of the terminal descent guidance system alleviates this problem by imposing constraints on the utilization of the guidance law during the gravity-turn mode.

Proposed Terminal Descent Guidance System Mechanization

During the gravity-turn descent, the guidance system modulates the thrust developed by the retro engines as necessary to effect a soft landing. Modulation of the thrust acceleration is accomplished by comparing the instantaneous acceleration required to soft-land, computed from Eq. (45), with the output of an on-board axial accelerometer and using the resultant error to adjust the differentially throttlable engines. A computer that evaluates A_T and has inputs from the on-board radar system and an output to an error amplifier is required to mechanize the guidance system. A simplified block diagram of the closed-loop terminal descent system, illustrating the guidance system integration, is shown in Fig. 14.

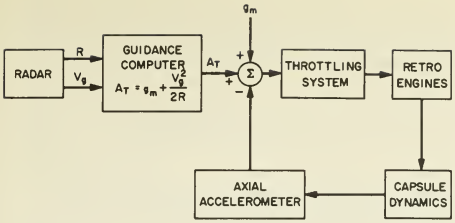


Fig. 14. Simplified block diagram of closed-loop terminal-descent system

A computer simulation of the vertical-plane descent was developed to evaluate the feasibility of using this guidance law to control the gravity-turn descent. It was assumed in the evaluation that all systems were ideal and that the modes required in this mechanization were not perturbed by imperfect executions. Using the ignition conditions of Table 5, preliminary results indicated that excessive amounts of fuel were required for landings in the VM-10 atmosphere. The high-altitude, low-velocity ignition condition for VM-10 results in a total velocity to be removed by the retro engines, the sum of the velocity acquired due to gravity and the surface-referenced velocity, which is greater than the total velocity for VM-7. The consequence of such an occurrence is that the fuel requirements for VM-10 are greater than for VM-7 where the ignition velocity is approximately a factor of two greater.

Two alternatives are available to prevent the fuel requirements from being dictated by the most dense atmosphere in which the maximum ballistic velocity reduction occurs. The obvious alternative is to select an ignition criterion, other than igniting at a given slant range, for which the ignition altitude decreases as the atmospheric density increases. However, because of the mechanization simplicity and the fact that preignition velocity and slant-range data are not required, the second alternative will be considered for reducing fuel requirements in the more dense atmospheres and, therefore, allow propulsion system sizing on the basis of the highest ignition velocity. The second alternative consists of introducing a mode in the guidance system which, for the denser atmospheres, allows the capsule to descend to an altitude-velocity condition comparable with conditions in less dense atmospheres. The descent to such a condition is accomplished by throttling the engines to some minimum thrust level, sufficient to maintain attitude control capability, at the completion of the stabilization and acquisition mode. During this minimum thrust period or drop mode, the altitude reduction and a small velocity reduction is achieved at the expense of a minimal amount of fuel and the atmospheric drag acting on the capsule.

The difficulty of implementing the drop mode arises in establishing a criterion for when the mode should be used and, if it is used, in defining its duration. The duration bounds can be determined by computing the capsule acceleration characteristics from the guidance law of Eq. (45). A typical set of acceleration characteristics, plotted as a function of slant range, is shown in Fig. 15. From the figure, it is observed that the duration is bounded by the VM-7 and VM-10 characteristics.

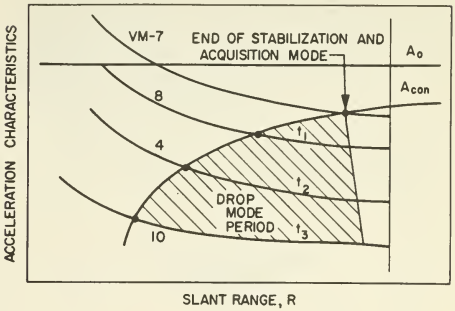


Fig. 15. Typical acceleration characteristics which illustrate duration of the drop mode

If a continuous acceleration contour can be defined as an inverse function of slant range, a drop mode can be defined whose duration and utilization is proportional to the atmospheric density for any atmosphere between the VM-7 and VM-10 bounds. Such a contour is given in Fig. 15 as A_{con}. Mathematically, it may be described as

$$A_{con} = A_0 - \frac{\alpha_K}{2R} \tag{46}$$

where A_0 is an asymptote of A_{con} and whose value is determined from the maximum thrust acceleration capability of the engines, and α_K is a constant which shapes the contour. The slant range value at the end of the stabilization and acquisition mode, A_{con}, and the VM-10 acceleration characteristic are the boundaries of the drop mode.

To establish the drop-mode guidance policy, the following criterion may be stated:

If, at the end of the stabilization and acquisition mode the value of A_{con} is greater than the computed value of A_T, begin a gravity-turn drop-mode descent. Continue in this state, continuously comparing A_{con} with A_T, until A_T equals or is greater than A_{con} or

$$A_T \geq g_m + \frac{V_g^2 + \alpha_K}{2R} \tag{47}$$

When this condition occurs, switch the thrust acceleration control to A_T.

The mechanization required to execute this guidance policy, which consists of a computer and sequencer, is illustrated in Fig. 16. The constants g_m, α_K, A_0, A_1, and A_2, where A_1 is the thrust acceleration required for stabilization and acquisition and A_2 is the minimum thrust acceleration during the drop mode, are placed in data

storage in the computer. The guidance computer and sequencer operations may be briefly described as follows:

The RMR marking signal activates a start mechanism and A_T is set equal to A_1. The term A_1 is sent to the output block where it is diverted to the thrust control loop and to the "If Doppler Reliable" block which denotes that the radar has acquired. When the doppler reliable signal occurs, computation of A_T begins. The term A_T is then compared with A_0 and, if $A_T < A_0$, the drop mode begins and A_T is set equal to A_2. The capsule continues to descend in the drop state until $A_T \geq A_0$, at which time a_K and A_0 are set to zero. The computed descent state is maintained until landing or some terminal mode is initiated.

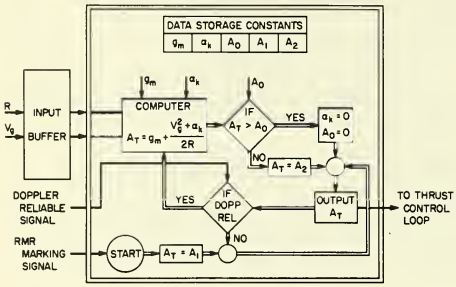


Fig. 16. Guidance computer and sequencer

Guidance System Perturbations

Dispersions in the computed thrust acceleration resulting from radar measurement errors in slant range and velocity can result in additional fuel requirements and severely affect the landing conditions. Random measurement noise predominantly affects the landing conditions; dispersions about the nominal zero velocity, zero altitude condition result. The total amount of fuel required remains approximately constant provided the expected value of the noise is zero. Constant measurement biases influence both the amount of fuel required and the landing conditions. The effect of a bias is to introduce a change in the computed thrust acceleration proportional to the magnitude of slant range and velocity. More fuel is required and a zero velocity condition is achieved above the surface if a bias increases the thrust acceleration; the converse also is true. Methods by which these adverse effects can be suppressed include filtering and the use of a terminal mode in the final near-surface segment of the descent.

Another phenomenon which must be considered is the terrain bias effect. Of particular interest is the constant-sloping terrain. Consider the geometry of the descent as shown in Fig. 17, where the surface slope is denoted by μ. From the geometry, ΔR is the difference between the measured slant range and the value of slant range as computed from the true altitude and path angle of the capsule. The dependence of ΔR on μ is determined from the expression

$$\Delta R = R \cot \gamma_g \tan \mu \qquad (48)$$

To accommodate this effect in the guidance law, the expression can be written as

$$A_T = g_m + \frac{V_g^2}{2R(1 + \cot \gamma_g \tan \mu)} \qquad (49)$$

The effect is simply an addition or subtraction of a proportional bias; the bias being additive for an "uphill" slope (μ negative) and subtractive for a "downhill" slope (μ positive).

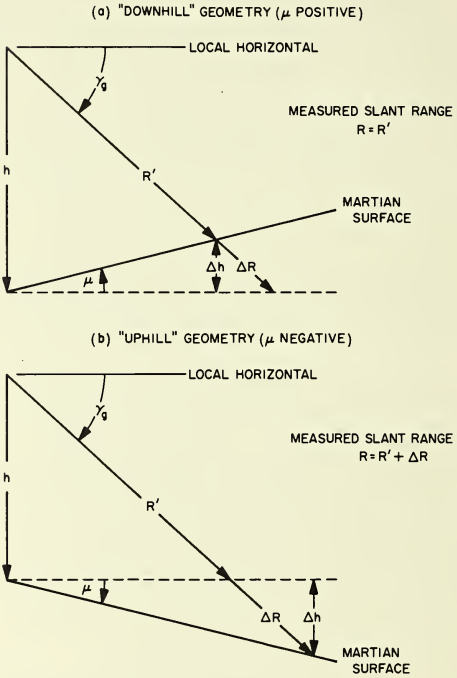


Fig. 17. Surface slope geometry

Simulated Flight Results and Performance Summary

The feasibility of the terminal descent guidance system concept can be demonstrated from simulated data describing the retro-controlled descent phase. The above-mentioned simulation of the vertical-plane descent equations of motion was modified to include the drop mode and used to obtain the simulated data. Descent characteristics of particular interest included the altitude-velocity and fuel consumption-retro descent time profiles.

For the purpose of feasibility evaluation, two types of simulated descents were defined. First, an "ideal descent" was defined where the capsule was descending in an unperturbed state over

perfectly smooth terrain. The results from this type of descent were primarily for demonstrating the drop-mode feasibility and illustrating the sensitivity of the descent characteristics to changes in retro ignition conditions. The second simulated descent type was defined as a "typical descent." For this descent type, the results were indicative of the effects of steady-state winds, measurement noise and biases, and terrain slopes on the descent characteristics.

The input data to the simulation were the atmospheric constants and the retro ignition conditions. The other program constants necessary to specify a particular descent included a stabilization and acquisition mode duration, assumed to be 3.5 sec, and the computer data storage constants, which were assumed to have the following values:

$$a_K = [1.976(0.494)3.458] \times 10^5 \text{ ft}^2/\text{sec}^2$$

$$A_0 = 4g_m$$

$$A_1 = 2g_m$$

$$A_2 = 0.8g_m$$

Shown in Figs. 18 and 19 are plots of the altitude versus surface-referenced velocity and fuel consumption versus retro descent time for the ignition conditions given in Table 5. The data shown are for ideal landings and, therefore, do not include steady-state wind or measurement errors. In each of the plots, the three retro-controlled descent modes are indicated; the duration of each mode can be computed from Fig. 19. From the figures, it is observed that, with the drop mode, the maximum amount of fuel is consumed by the VM-7 atmosphere. Previous studies, using the same ignition conditions and without the drop mode, have shown that the VM-10 atmosphere required approximately from 20 to 30% more fuel than VM-7 and resulted in a descent time exceeding 100 sec.

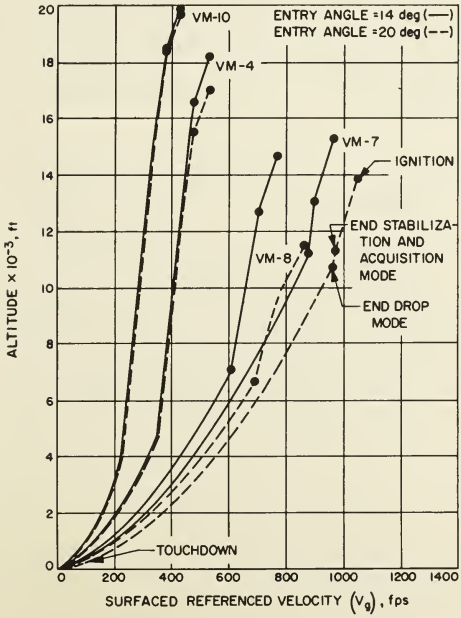


Fig. 18. Ideal landings—altitude-velocity profile

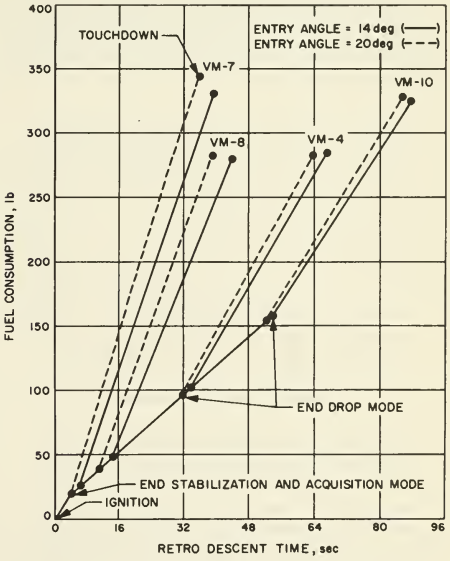


Fig. 19. Ideal landings—fuel consumption-retro descent time characteristics

The selection of a value for the shaping constant, a_K, is facilitated by the parametric data shown in Fig. 20. In the figure, the altitude at which the computed descent is initiated (dotted lines) and fuel consumption characteristics are plotted for each of the atmospheric models and for the shaping constant range indicated above. The range of values for a_K was determined by considering the fuel requirements for the VM-10 atmosphere and the thrust acceleration required at the initiation of the computed descent. If a_K shapes the contour such that the computed descent initial altitude is too near the surface, then the thrust acceleration needed to soft-land may exceed the retro-engine capability. On the basis of the data shown, the value of a_K selected was 2.47×10^5 ft^2/sec^2. (The previous data and the remaining data were computed using this particular value.)

To evaluate, on a preliminary basis, the effect of measurement errors, random noise with zero mean, and constant 1-σ, values of 4 ft/sec and 6 ft were added to the measured values of velocity and

slant range, respectively. (Altitude scaling of the noise was not considered in the measurement error model assumed.) In addition, constant biases of magnitude equal to 1% of the values of slant range and velocity at the completion of the drop mode were also included. Descents were simulated for each of the ignition conditions listed in Table 5. The results, in all cases, showed that the trajectory became extremely erratic as the capsule approached the surface; significant velocity-altitude changes occurred because of the decreasing signal-to-noise ratio. The amount of fuel consumed for these descents remained approximately the same as for the ideal landing cases. These results, in general, demonstrated that measurement errors can severely affect landing conditions and that filtering techniques and terminal modes must be considered in the terminal descent system design.

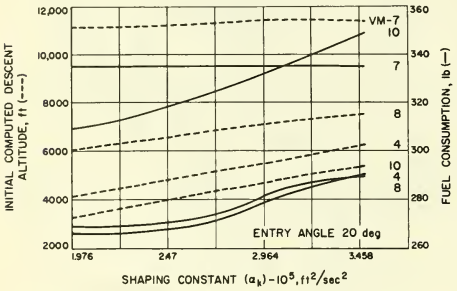


Fig. 20. Ideal landings—effect of shaping constant variation on fuel consumption and initial computed descent altitude

The effect of a constant surface slope on fuel consumption and retro descent time is shown in Fig. 21 for a downhill slope. The ignition condition chosen to illustrate this effect was the case of the VM-8, 14-deg entry angle. As shown in the figure, considerable increase, in both the amount of fuel required and descent time of 14 and 31%, respectively, occurred for the 20-deg slope change. In an uphill slope, the fuel consumption and descent time are less. Near the surface, the thrust acceleration increases in an attempt to remove the final increment of velocity which has accumulated as a result of the initially low-thrust acceleration. (The increase in slant range for the uphill slope decreases the thrust acceleration.) The value of the thrust acceleration required may be of sufficient magnitude to exceed the throttling capability of the retro engines. Thus, descents over sloping terrain may represent significant problems from the fuel consumption and throttling capability points-of-view; however, analyses using more realistic terrain models are required to determine the total magnitude of the effect.

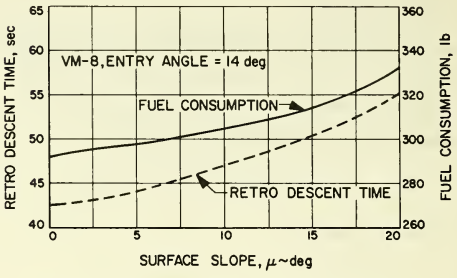


Fig. 21. Typical landings—effect of surface slope on descent time and fuel consumption

In summary, the preliminary results discussed in this section indicate that the proposed guidance system appears feasible. The inclusion of the drop mode as a terminal descent mode eliminates to a great extent the dependence of the descent system design on the relatively unknown Martian atmosphere. Further effort, which concentrates on particular problem areas relevant to the guidance system, is required before the information is available for conducting the necessary tradeoff studies to finalize a guidance system design.

IV. References

(1) "Proposed Biological Exploration of Mars Between 1969 and 1973," Nature, pp. 974-980, June 5, 1965.

(2) Thompson, R. P. "The Practical Problem of Landing on Mars," Aeronautics and Astronautics, pp. 66-73, July 1966.

(3) Stone, I., "Atmosphere Data to Alter Voyager Design," Aviation Week and Space Technology, pp. 66-69, November 22, 1965.

(4) "Altimeter Study for the Ranger Landing System," Aeronutronics Final Report V-2252, Vol. II, August 30, 1963.

-- NOTES --

-- NOTES --

No. 67-620

ITERATIVE GUIDANCE APPLIED TO GENERALIZED MISSIONS

by

HELMUT J. HORN, DORIS C. CHANDLER

and

VOLIS L. BUCKELEW

NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-620

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

1.03, 2.06, 8.02

ITERATIVE GUIDANCE APPLIED TO GENERALIZED MISSIONS

Helmut J. Horn, Doris C. Chandler, and Volis L. Buckelew
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

The iterative guidance mode (IGM) was developed to satisfy the launch vehicle requirements of the Saturn-Apollo flights. Improvements in the formulation of the guidance equations extend the application of the IGM to more general missions: (1) A more sophisticated prediction of the "effective" gravity vector which permits use of the guidance scheme for extended trajectory arcs without deterioration of the optimization characteristics; (2) development of equations which permit the use of total energy as an additional end constraint, reducing or entirely eliminating the need of precomputation and curve-fitting for a major group of missions; (3) the reduction of sensitivity to noise in the acceleration measurement, which is achieved by computing and filtering a fictitious thrust, the product of measured acceleration, and computed mass. The effectiveness of these modifications, which are presented against the background of a short derivation of the basic equations, is numerically evaluated for selected trajectories.

Nomenclature

Symbol	Definition
E	energy
F	thrust force
g	gravitation
g_o	gravitation at r_o
R	value to be optimized
r	radius vector
r_o	earth radius
t	time
T	time to go
v	velocity
K, K_1, K_2	constants
L,J,P	substitutions
φ	range angle (variable)
Φ	total range angle
τ	time to complete consumption
ω	range angle rate
χ	thrust (vehicle) attitude
θ	path angle
Δ	difference

Subscripts

i	initial
T	terminal, end value
m	mean
F	caused by thrust force
x,y	components

Superscripts

$\rightarrow$	vector
$\sim$	applies to fixed attitude
*	effective
','	identifications

I. Introduction

The implementation of the iterative guidance mode (IGM) for the Saturn-Apollo flights resulted in a set of simplified equations optimized for the quick solution of this specified guidance problem. However, the exploration of the inherent limitations of the system and the development of more sophisticated formulations which permit application to more general missions were neglected.

The Apollo Applications Program has increased interest for a more general development of the IGM. A brief derivation of the two-dimensional IGM equations will provide the background to analyze and extend the range of validity of the system.

II. Derivation of Two-Dimensional IGM Equations and Identification of Limitations

The IGM is based on Lawden's[1] analytical solution of the power-flight optimization:

$$\tan \chi = \frac{\partial R/\partial \dot{y} + (T-t)\ \partial R/\partial y}{\partial R/\partial \dot{x} + (T-t)\ \partial R/\partial x} . \qquad (1)$$

If only a specific terminal velocity vector without position constraints is required, the partial derivatives with respect to the position coordinates are set equal to zero, and equation (1) becomes:

$$\tan \tilde{\chi} = \frac{\partial R/\partial \dot{y}}{\partial R/\partial \dot{x}} = \text{constant}, \qquad (2)$$

indicating a constant thrust attitude as optimum solution[2,3]. This case is very simply illustrated in figure 1, where v_1 is the initial velocity, v_g the velocity increment resulting from the gravitation field, and the sum of the thrust caused velocity increments is shown as the broken line $P_3 P_4'$.

For a given total impulse, the maximum velocity in the prescribed direction v_T is reached by stretching this broken line into the straight line $P_3 P_4''$, and P_4''' gives the absolute maximum of the achievable velocity if the direction is free. To satisfy one position constraint, a degree of freedom must be restored by re-introducing the "y" derivative into equation (2), and rewriting it as

$$\tan \chi = \tan \tilde{\chi} - (K_1 - K_2)t. \qquad (3)$$

By treating $(K_1 - K_2 t)$ as a small term, a closed form or an iterative solution for the thrust angle χ

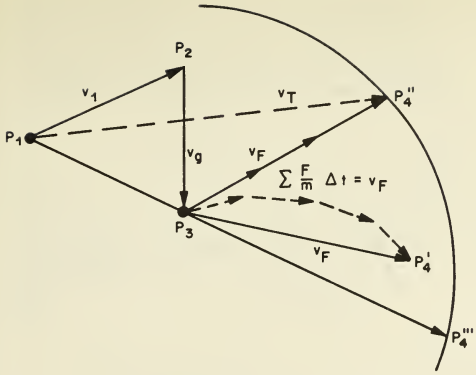


Figure 1. Velocity Hodograph

as function of the known vehicle parameters and the initial and final state vector can be calculated. Equation (1) is valid if the integrated effect of the gravitation field is a function of flight time only and is independent of the flight path, e.g., in a uniform gravitation field. This poses two problems: the prediction of an "effective" gravity vector

$$\vec{g}^* = \frac{1}{T}\int_o^T \vec{g}\, dt \qquad (4)$$

for the remainder of the flight and the sensitivity of the actual gravitation effects to changes of the flight path. For typical Saturn-Apollo missions with central angles approaching 30°, the mean of the instantaneous and end point gravitation vectors proved to be an adequate approximation for the effective gravity, and the resulting payload losses were negligible. This paper explores more sophisticated "effective" gravity predictions.

A second critical point in using the IGM is the correct definition of the end conditions. If only the velocity vector is prescribed, equation (2) gives a ready answer. If a position constraint normal to the velocity direction is added, equation (3) provides a good solution. A deviation from this flight geometry, i.e., a position constraint in a different direction, causes a high sensitivity to noise in the measured data, especially in the acceleration measurement. Where it is not possible to reformulate the end conditions to achieve this favorable geometry, an improved filtering scheme is given.

Another set of end conditions defining the end-velocity as function of the vehicle position requires some changes in the existing IGM equations. A solution for the most important case of this type, optimization of flight for a given total energy, is described in Section IV.

III. Definition of the "Effective" Gravity Vector

Since the gravitation vector is known as a function of the vehicle position, and the IGM predicts the path of the vehicle to its end position at each iteration step, a numerical integration of the gravitation is possible. Because of the excessive computation effort, several approximations are derived and evaluated.

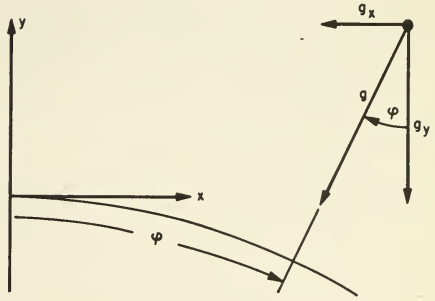


Figure 2. Gravitation Vector

The components of the effective gravitation vector g* can be expressed as the product of a function of the instantaneous gravity direction and a "mean" value of the gravity magnitude g' (figure 2):

$$g_x^* = \left(\frac{1}{T}\int_o^T \sin\varphi\, dt\right) g' \qquad (5)$$

$$g_y^* = \left(\frac{1}{T}\int_o^T \cos\varphi\, dt\right) g'. \qquad (6)$$

The central angle φ is approximated by assuming a constant angular acceleration:

$$\varphi = \omega_m t + \frac{1}{2}\dot{\omega}t(t - T). \qquad (7)$$

Closed form integration of the gravity components is possible if the variation against a time linear central angle $1/2\ \dot{\omega}t(t - T)$ is considered as a small value.

$$\sin\varphi \approx \sin\omega_m t + \frac{1}{2}\dot{\omega}t\,(t - T)\cos\omega_m t \qquad (8)$$

$$\cos\varphi \approx \cos\omega_m t - \frac{1}{2}\dot{\omega}t\,(t - T)\sin\omega_m t \qquad (9)$$

$$\int \sin\varphi\, dt = -\frac{1}{\omega_m}\cos\omega_m t - \frac{1}{2}\dot{\omega}T\left(\frac{t}{\omega_m}\sin\omega_m t + \frac{1}{\omega_m^2}\cos\omega_m t\right) + \frac{1}{2}\dot{\omega}\left[\left(\frac{t^2}{\omega_m} - \frac{2}{\omega_m^3}\right)\sin\omega_m t + \frac{2t}{\omega_m^2}\cos\omega_m t\right] + C. \qquad (10)$$

By introducing the limits and substituting

$$\text{for } t_i = \varphi_i = 0$$

$$\int \sin \omega_m \, dt = 0$$

$$\text{and } t = T;\ \varphi(T) = \emptyset = \omega_m T;\ \dot{\omega} T = \Delta\omega$$

$$\frac{1}{T}\int_0^T \sin\varphi \, dt = -\frac{1}{\emptyset}\left[\left(1 - \frac{1}{2}\frac{\Delta\omega}{\omega_m}\right)\cos\emptyset + \frac{1}{\emptyset}\frac{\Delta\omega}{\omega_m}\sin\emptyset - \left(1 + \frac{1}{2}\frac{\Delta\omega}{\omega_m}\right)\right]. \tag{11}$$

Substituting again yields

$$\cos\emptyset = 1 - 2\sin^2(\emptyset/2) \tag{12a}$$

$$\sin\emptyset = 2\sin\emptyset/2\cos\emptyset/2 \tag{12b}$$

$$\frac{1}{T}\int_0^T \sin\varphi \, dt = \left[\left(1 - \frac{1}{2}\frac{\Delta\omega}{\omega_m}\right)\sin\frac{\emptyset}{2} - \frac{1}{\emptyset}\frac{\Delta\omega}{\omega_m}\cos\frac{\emptyset}{2}\right]\frac{\sin\emptyset/2}{\emptyset/2} + \frac{1}{\emptyset}\frac{\Delta\omega}{\omega_m}. \tag{13}$$

In a similar way, the second component can be calculated as

$$\frac{1}{T}\int_0^T \cos\varphi \, dt = \left[\left(1 - \frac{1}{2}\frac{\Delta\omega}{\omega_m}\right)\cos\frac{\emptyset}{2} + \frac{1}{\emptyset}\frac{\Delta\omega}{\omega_m}\sin\frac{\emptyset}{2}\right]\frac{\sin\emptyset/2}{\emptyset/2}. \tag{14}$$

For the special case of an unpropelled satellite in circular orbit, after substituting $\Delta\omega = 0$ and ang $g' = g$, the equations (5) and (6) degenerate into

$$g_x^*/g = \sin\frac{\emptyset}{2}\frac{\sin\emptyset/2}{\emptyset/2} \tag{15}$$

$$g_y^*/g = \cos\frac{\emptyset}{2}\frac{\sin\emptyset/2}{\emptyset/2} \tag{16}$$

$$g^* = g\frac{\sin\emptyset/2}{\emptyset/2}, \qquad \emptyset^* = \emptyset/2. \tag{17}$$

At this point, it should be noticed that the single integral to velocity and the double integral to position result in two different definitions of the "effective" gravitation:

$$\vec{g}_V^* = \frac{1}{T}\int_0^T \vec{g}\, dt \tag{18a}$$

and

$$\vec{g}_R^* = \frac{2}{T^2}\int_0^T\int \vec{g}\, dt^2. \tag{18b}$$

By integration of equation (18b), the direction-dependent components of the "position effective" gravitation vector are found as

$$g_{R\cdot x}^* = g''(2/T^2)\int_0^T\int \sin\varphi \, dt^2 = \frac{2g''}{\emptyset}\left[\left(1 + \frac{1}{2}\frac{\Delta\omega}{\omega_m}\right) - \frac{\sin\emptyset/2}{\emptyset/2}\left(\left(1 - \frac{\Delta\omega}{\omega_m}\right)\cos\frac{\emptyset}{2} - \frac{3}{\emptyset}\frac{\Delta\omega}{\omega_m}\sin\frac{\emptyset}{2}\right)\right] \tag{19}$$

and

$$g_{R\cdot y}^* = g''(2/T^2)\int_0^T\int \cos\varphi \, dt^2 = \frac{2g''}{\emptyset}\left[\frac{3}{\emptyset}\frac{\Delta\omega}{\omega_m} + \frac{\sin\emptyset/2}{\emptyset/2}\left(\left(1 - \frac{\Delta\omega}{\omega_m}\right)\sin\frac{\emptyset}{2} - \frac{3}{\emptyset}\frac{\Delta\omega}{\omega_m}\cos\frac{\emptyset}{2}\right)\right]. \tag{20}$$

These equations for effective gravity components as functions of the range angle were combined with approximations for the magnitude of the gravitation, leading to the following definitions:

Case 1a: $$g_V'' = g_R'' = \frac{1}{2}(g_i + g_T) \tag{21}$$

where $g = -\mu/r^2$.

Case 1b: $$g_V'' = g_R'' = \frac{1}{4}(g_i + 2g_{(T/2)} + g_T). \tag{22}$$

A better approximation for the position part would have been

$$g_R'' = \frac{1}{16}(7g_i + 8g_{(T/2)} + g_T), \tag{23}$$

but since the position constraint is less critical than the velocity constraint, this was not pursued further.

Although case 1c uses the same approximation as case 1a, it breaks the trajectory arc into two sub-arcs.

Case 1c: $$g_{V\cdot x}^* = \frac{1}{2T}\left[\left(g_i + g_{(T/2)}\right)\int_0^{T/2} \sin\varphi \, dt + \left(g_{(T/2)} + g_T\right)\int_{T/2}^T \sin\varphi \, dt\right] \tag{24}$$

and the similar equations for the y-component and the double integrals. Next, the effective g was approximated as a simultaneous function of magnitude and direction of the actual gravity, yielding case 2,

$$\underline{\text{Case 2}}: \quad g = g_i + \dot{g}t \tag{25a}$$

$$\varphi = \omega_m t + \frac{1}{2}\dot{\omega}(t - T) \tag{25b}$$

and case 3, which is identical to case 2, but breaks the trajectory arc into two sub-arcs as in case 1c.

These approximations, together with a case 0 representing the existing IGM as reference, were applied to four missions:

A. a typical Saturn V ascent into parking orbit,

B. a translunar injection out of parking orbit,

C. the same, but with reduced thrust,

D. a direct ascent to translunar injection.

The nominal values of the range angle and the errors and nominal values of the direction ϕ^* and magnitude g* of the effective gravity for velocity and position are shown in table 1.

Mission	Nominal		Error in ϕ_V^* (°)					
	ϕ(°)	ϕ_V^*(°)	Case 0	Case 1a	Case 1b	Case 1c	Case 2	Case 3
A	25.8	10.4	1.45	0.13	0.13	0.06	0.10	0.06
B	23.6	11.1	0.47	0.08	0.08	0.03	0.05	0.03
C	93.1	43.7	9.25	2.15	2.15	0.69	1.14	0.10
D	50.9	20.2	3.01	0.32	0.32	0.09	0.51	0.14

Mission	Nominal		Error in g_V^* (m/s^2)					
	ϕ(°)	$g_V^*(\frac{m}{s^2})$	Case 0	Case 1a	Case 1b	Case 1c	Case 2	Case 3
A	25.8	9.20	0.06	0.08	0.05	0.04	0.07	0.04
B	23.6	9.17	0.14	0.06	0.02	0.02	0.06	0.02
C	93.1	7.33	1.54	0.51	0.10	0.15	0.50	0.13
D	50.9	9.13	0.71	0.16	0.09	0.09	0.09	0.09

Mission	Nominal		Error in ϕ_R^* (°)					
	ϕ	ϕ_R^*	Case 0	Case 1a	Case 1b	Case 1c	Case 2	Case 3
A	25.8	6.1	5.13	0.12	0.12	0.01	0.10	0.01
B	23.6	7.2	3.56	0.06	0.06	0.02	0.04	0.02
C	93.1	29.5	2.80	0.79	0.79	0.15	1.18	0.12
D	50.9	11.7	10.32	0.43	0.43	0.15	0.10	0.10

Mission	Nominal		Error in g_R^* (m/s^2)					
	ϕ	g_R^*	Case 0	Case 1a	Case 1b	Case 1c	Case 2	Case 3
A	25.8	9.27	0.09	0.06	0.05	0.04	0.08	0.04
B	23.6	9.22	0.18	0.08	0.05	0.02	0.05	0.03
C	93.1	8.19	2.31	1.11	0.22	0.29	0.52	0.15
D	50.9	9.32	0.87	0.29	0.09	0.04	0.19	0.03

Table 1. "Effective" Gravitation Prediction

The improvement of all approximations over the present IGM is obvious in all cases except where the error is already in the noise level, about .1° and .1 m/s^2. The pattern generally verifies that the use of additional information improves the results, showing the least improvement for case 1a, and the most improvement for case 3. The rating of the three intermediate cases depends on the specific mission and on the weight assigned to the error components.

IV. Trajectory Optimization for Total Energy

Similar to the optimization for a prescribed end velocity vector, which was solved by equation (2), the optimization for a prescribed total energy and the velocity direction has two constraints, leaving one degree of freedom; i.e., one parameter must be determined. Starting with equation (1), we rotate the coordinate system so that the y-axis is the local vertical at the injection point. With the x-coordinate horizontal, the x_T displacement will not cause a first order change of the energy; i.e., we have no x_T-constraint, and optimization requires that the partial derivative with respect to x equal zero. Equation (1) can now be rewritten as

$$\tan \chi = \frac{\partial R/\partial \dot{y}}{\partial R/\partial \dot{x}} + (T - t)\,\frac{\partial R/\partial y}{\partial R/\partial \dot{x}}\,, \tag{26}$$

indicating a time-linear steering law.

The hodograph in figure 1 shows that the velocity gain for a slightly curved $\int a\,dt$, the distance $P_3 P_4'$, is shorter only by a second-order term than the straight line $P_3 P_4''$. However, the integral shows a first-order change in displacement.

Similarly, as in the original derivation for the position constraint, we first calculate the angle $\tilde{\chi}$, which satisfies the end-velocity, using an estimated cutoff altitude for the potential energy computation, and vary it by a small linear term

$$\chi = \tilde{\chi} - K(t - \tfrac{1}{2} T). \tag{27}$$

The small angle approximation using second order terms yields

$$\cos \chi = \left[1 - \tfrac{1}{2} K^2 (t - \tfrac{1}{2} T)^2\right] \cos \tilde{\chi} + K(t - \tfrac{1}{2} T) \sin \tilde{\chi} \tag{28}$$

and

$$\sin \chi = \left[1 - \tfrac{1}{2} K^2 (t - \tfrac{1}{2} T)^2\right] \sin \tilde{\chi} - K(t - \tfrac{1}{2} T) \cos \tilde{\chi}. \tag{29}$$

Applying the rocket equation,

$$F/m = V_{ex}/(\tau - t), \tag{30}$$

to the acceleration components

$$\ddot{x}_F = (F/m) \cos \chi \tag{31}$$

$$\ddot{y}_F = (F/m) \sin \chi \tag{32}$$

and substituting

$$L_1 = \int \frac{1}{\tau - t}\, dt = \ln \frac{\tau}{\tau - t} \tag{33}$$

$$L_2 = \iint \frac{1}{\tau - t}\, dt^2 = -(\tau - t)\, L_1 + t \tag{34}$$

$$J_1 = \int \frac{t}{\tau - t}\, dt = \tau L_1 - t \tag{35}$$

$$J_2 = \iint \frac{t}{\tau - t}\, dt^2 = -\tau \left[(\tau - t)\, L_1 - t\right] - \tfrac{1}{2} t^2 \tag{36}$$

$$P_1 = \int \frac{t^2}{\tau - t}\, dt = \tau\,[\tau L_1 - t] - \tfrac{1}{2} t^2 \tag{37}$$

$$P_2 = \iint \frac{t^2}{\tau - t}\, dt^2 = -\tau \left\{\tau[(\tau - t)L_1 - t] + \tfrac{1}{2} t^2\right\} - \tfrac{1}{6} t^3, \tag{38}$$

we get the velocity components caused by the thrust,

$$\dot{x}_{FT} = \dot{\tilde{x}}_{FT} + \dot{\tilde{y}}_{FT}\left[(\tau - \tfrac{1}{2} T) - \frac{T}{L_1}\right] K - \dot{\tilde{x}}_{FT}\, \tfrac{1}{2}\left[(\tau - \tfrac{1}{2} T)^2 - \frac{(\tau - \tfrac{1}{2} T)\, T}{L_1}\right] K^2 \tag{39}$$

$$\dot{y}_{FT} = \dot{\tilde{y}}_{FT} - \dot{\tilde{x}}_{FT}\left[(\tau - \tfrac{1}{2} T) - \frac{T}{L_1}\right] K - \dot{\tilde{y}}_{FT}\, \tfrac{1}{2}\left[(\tau - \tfrac{1}{2} T)^2 - \frac{(\tau - \tfrac{1}{2} T)\, T}{L_1}\right] K^2, \tag{40}$$

and the altitude term

$$y_{FT} = \tilde{y}_{FT} - \tilde{x}_{FT}\left[(\tau - \tfrac{1}{2} T) - \tfrac{1}{2}\frac{T^2}{L_2}\right] K - \tfrac{1}{2}\,\tilde{y}_{FT}\left[(\tau - \tfrac{1}{2} T)^2 - (\tfrac{1}{4}\tau - \tfrac{1}{6} T)\,\frac{T^2}{L_2}\right] K^2. \tag{41}$$

The total energy is

$$E = \tfrac{1}{2} v^2 - \frac{g_o r_o^2}{r}, \tag{42}$$

and its derivative, which has to equal zero for an optimum solution, becomes

$$dE_T = v_T\, dv + g_o\, (r_o/r)^2\, dr = 0, \tag{43}$$

where $v = \dot{x}_T \cos \theta_T + \dot{y}_T \sin \theta_T$. (43a)

The thrust generated portions, $d\dot{x}_{F,T}$, $d\dot{y}_{F,T}$ and $dy_{F,T}$ can be substituted for the complete values $d\dot{x}_T$, $d\dot{y}_T$ and dy_T because the initial conditions and the gravitation effects are independent of K. Also, θ_T is prescribed and constant. Therefore,

$$dv = \left(\frac{\partial \dot{x}_T}{\partial K} \cos \theta_T + \frac{\partial \dot{y}_T}{\partial K} \sin \theta_T\right) dK \qquad (44)$$

and

$$dr = dy = (\partial y/\partial K)\, dK, \qquad (45)$$

leading to the equation

$$K = \frac{v_T \sin(\tilde{\chi}-\theta_T)(\vartheta L_1 - T) + g_T \cos \tilde{\chi}(\tau - T)(\vartheta L_1 - T)}{v_T \cos(\tilde{\chi}-\theta_T)\vartheta(\vartheta L_1 - T) - g_T \sin \tilde{\chi}\left[\vartheta^2(\tau - T)L_1 - \vartheta^2 T + \left(\frac{1}{2}\tau - \frac{1}{3}T\right)T^2\right]}, \qquad (46)$$

where

$$\vartheta = \tau - 1/2\ T.$$

It is interesting to apply this equation to two special cases: the injection into a circular orbit and the injection at perigee into a parabolic orbit, assuming initial conditions requiring no vertical thrust component,

$$\tilde{\chi} = \theta_T = 0, \qquad (47)$$

and for the circular orbit,

$$v_T\, \omega_T = g_T \qquad (48)$$

$$K = \frac{g_T(\tau - T)}{v_T(\tau - \frac{1}{2}T)} = \omega_T \frac{1 - \frac{T}{\tau}}{1 - \frac{1}{2}\frac{T}{\tau}} \leq \omega_T. \qquad (49)$$

This means that the angular rate is lower than the orbital rate, and the orbital rate is approached if T becomes very small compared to τ. Physically, this means a mass ratio approaching unity. This is the case if either the velocity-to-be-gained is very low, or if the exhaust velocity (or the specific impulse) is very high.

For the escape, or parabolic case, by the same calculation, the angular tilt rate K is equal to or below one-half the parabolic orbital rate at perigee, or equal to or below $1/\sqrt{2}$ times the circular orbital rate. The tilt rate K from equation (45) was applied to a typical planetary injection with range angle of approximately 30° (table 2a) and to a lunar injection with range angle of 23° (table 2b). For comparison, a trajectory was computed with the existing equations, using the radius and velocity resulting for the first trajectory as end constraint. In addition, a trajectory was computed, using the $\tilde{\chi}$ for equation (2). Initially, the end velocity from the first, energy-optimized trajectory was used as constraint, but the propulsion kept working until the required energy was reached. The $\tilde{\chi}$ trajectory showed the poorest performance. As to be expected, the velocity optimization led to too low an

Guidance Mode	Radius (km)	Velocity (m/s)	Energy (km^2/s^2)	Inj. Mass (lbs)
Energy	-	-	18.066	108,730
Pos. & Veloc.	6711.9	11697.5	-	108,676
$\tilde{\chi}$	6682.8	11720.2	18.068	108,456
Nominal Values	6712	11697.5	18.066	-

Table 2a. Typical Planetary Injection

Guidance Mode	Radius (km)	Velocity (m/s)	Energy (km^2/s^2)	Inj. Mass (lbs)
Energy	-	-	-4.590	133,241
Pos. & Veloc.	6642.3	10,744	-	133,225
$\tilde{\chi}$	6628.8	10,755	-4.594	133,024

Table 2b. Typical Lunar Injection

injection point, and the energy deficit had to be made up by increasing the cutoff velocity. The performance difference between the "energy" guidance and the "position-velocity" guidance is not quite so obvious. A detailed inspection of the trajectory print-out shows a rather constant turning rate for the energy guidance, while the position velocity trajectories have a tendency to go nose-up at the beginning of flight, compensated by a later nose-down attitude. The reason may be a higher sensitivity to errors caused by the simplifications of the guidance scheme if applied over extended trajectory arcs. Trajectories computed for higher injection energy, 23 km^2/sec^2, show the same trend.

The energy-guidance equations can therefore be considered valid for the forseeable missions.

V. Reduction of Noise Sensitivity

The major source of data noise is the measurement of the vehicle acceleration. Rather than using the highly nonlinear thrust acceleration, the reciprocal value, which ideally would be linear with time, has been programmed for the Saturn flights.

$$\frac{m}{F} = \frac{1}{F}(m_o + \dot{m}t). \qquad (50)$$

Extreme thrust fluctuation and a feedback through a propellant level sensor, propellant slosh, and the propulsion (propellant utilization) system caused marginal stability and started investigations for

better procedures. The best solution, wherever applicable, is the change of the flight geometry to use the position constraint normal to the injection velocity direction[4]. This, however, is not always permissible. Therefore, a different filtering technique was developed. The thrust has ideally a constant level, and filtering the thrust level would therefore be very efficient. Since no thrust measurement is available, the measured acceleration is multiplied with the calculated mass, the product is filtered and thereafter divided by the calculated mass[5]. A numerical analysis showed improved smoothness and reduced time lag of the result, compared with the filtering of the reciprocal acceleration. Perhaps more important is the absence of extreme overshoots in the commanded attitude toward the end of flight, which can make an early opening of the guidance loop necessary thereby reducing accuracy and optimization. At this time, no optimization of the filter and no quantitative analysis in depth has been made.

VI. Conclusions

The efforts to make the IGM applicable to a more general type of missions were concentrated on three specific points. First, better prediction of the gravitation effects permits a considerable expansion of the total range angle and improves at the same time the performance optimization. The second point, the energy optimization eliminates a considerable amount of precalculation and curve fitting. Once it is implemented, it will reduce the required numbers to be changed for a change in lunar landing site from 90 to about 60, possibly less. The desensitization against noise is, at this time, more a standby or a desirable improvement than an immediate necessity. The next phase will be to extend the shown principles to three dimensions, to apply them simultaneously rather than one at a time, and to implement and check them out on a flight computer. The proper balance between the desirable degree of sophistication and the necessary simplicity will play a decisive role in the dual formulation. Together with clearer definitions of future missions, it will influence the utilization of additional efforts, e.g., use of the more general formulation of the basic optimization equation[7].

$$\lambda_x = -(\partial R/\partial \dot{x}) \cos \omega_x(t-T) + (\partial R/\partial x)[\sin \omega_x(t-T)]/\omega_x \quad (51)$$

$$\lambda_y = -(\partial R/\partial \dot{y}) \cosh \omega_y(t-T) + (\partial R/\partial y)[\sinh \omega_y(t-T)]/\omega_y \quad (52)$$

and

$$\tan X = \lambda_y/\lambda_x \quad (53)$$

or the introduction of higher order terms into the presently used approximations.

References

1. Lawden, D. F., "Optimal Rocket Trajectories," Jet Propulsion 27, 1957, p. 1263.
2. Horn, H. J., "Application of an Interative Guidance Mode to a Lunar Landing," NASA TN D-2967.
3. Smith, I. E., "General Formulation of the Iterative Guidance Mode," NASA TM X-53414, Marshall Space Flight Center.
4. Chandler, D. C., H. J.Horn, R. S. Ryan, "Optimization Problems Caused by Interaction of Guidance and Propellant System." AAS Space Flight Mechanics Specialist Conference, U. of Denver, Colorado, July 1966, Prepring 66-117.
5. Buckelew, V. L., "Implementation of Methods to Reduce the Sensitivity of the Iterative Guidance Mode to Thrust Fluctuations of the AS-501 Vehicle," NASA TM X-53421, Marshall Space Flight Center.
6. Deaton, A. W., "Filtering of Acceleration Measurements for Guidance Calculations," to be published, Marshall Space Flight Center.
7. Fried, B. D., "Trajectory Optimization for Powered Flight in Two or Three Dimensions," Space Technology, Edited by H. Seifert, 1959, p. 4-9.

No. 67-621

A SIMPLIFIED EXPLICIT GUIDANCE SCHEME FOR ORBITAL INJECTION

by

HAROLD N. SCOFIELD
NASA Marshall Space Flight Center
Huntsville, Alabama

AIAA Paper
No. 67-621

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

2.06, 8.04, 10.06

-- NOTES --

A SIMPLIFIED EXPLICIT GUIDANCE SCHEME FOR ORBITAL INJECTION

Harold N. Scofield
Astrionics Laboratory
NASA, George C. Marshall Space Flight Center
Huntsville, Alabama

Abstract

Onboard computer equations for rocket vehicle guidance to elliptical orbit are developed. The guidance scheme computational requirements are held to the minimum consistent with accurate orbital injection, near optimum fuel comsumption, and an explicit formulation. Iteration is absent from the onboard computations; the pitch and yaw command angles are functions of the instantaneous state of the vehicle and two integrals evaluated by the onboard computer. Digital simulation results are shown for circular orbit injection with the Saturn V vehicle. Guidance scheme errors are less than one meter and one-tenth meter per second for disturbances caused by wind and variations in rocket engine parameters. Payload loss caused by the suboptimum guidance law is less than one-tenth percent of the gross weight delivered to orbit.

Symbols

f_0, f_1, f_2	coefficients for time polynomial used to decrease fuel consumption.
F	rocket thrust, N.
k	attitude command integral control law gain coefficient, rad/m/s.
m	rocket mass, kg.
r	scalar radius in XYZ coordinate system, m.
T	time from second stage ignition, s.
T_d	time remaining to orbital injection, s.
V	scalar velocity in XYZ coordinate system, m/s.
V_{ex}	velocity of rocket engine exhaust gas, m/s.
X, Y, Z	displacement coordinates in an earth-centered, space direction fixed coordinate system, m.
χ_p, χ_y	attitude command angles in pitch and yaw, rad.
τ	$\frac{V_{ex}}{F/m}$, s.
I_{sp}	specific impulse, s
Subscript c	command (or desired) value.
Subscript d	difference between target value and value at time T.
Subscript o	initial value
Subscript T	target value.

Introduction

The problem of steering a rocket vehicle from the earth's surface into elliptical orbit is considered. The resulting implementation is explicit; that is, the target orbit parameters appear directly in the formulation, and only a few constants dependent on the rocket performance are required. The fuel expenditure in reaching the target orbit is nearly optimum even though the constraints implied by flying a fuel minimum trajectory are not considered in the derivation of the steering law.

Unpublished analyses have shown that the first stage of a three stage Saturn class vehicle need not be guided in boost to low earth orbit. The first stage steering angles are functions of time only; therefore, the dispersions at first stage burnout are very large. However, an examination of the total flight shows such small degradations in fuel economy and orbital injection accuracy that these losses are difficult to determine. Since atmospheric forces are negligible in the two upper stages, the foregoing represents a significant simplification in the synthesis and analysis of the required guidance scheme.

The state of the vehicle motion is assumed to be measured in a Cartesian reference system fixed at the center of the earth with the X axis passing through the launch point, the Z axis pointing downrange parallel to the plane of the target orbit, and the Y axis completing a right-handed system. The XYZ system becomes space direction fixed at the instant of launch; i.e., it does not rotate with the earth after launch. Other quantities required from the vehicle navigation system are the radius in the XYZ system, r, the radius rate, $\dot{r}$, and the magnitude of the vehicle acceleration exclusive of gravitational forces. The guidance calculations result in pitch, χ_p, and yaw, χ_y, command attitude angles. Both angles are zero at launch and have sign definitions such that positive command angles steer the vehicle toward the +Z and +Y directions, respectively.

Fuel optimum flight dictates that injection into the target orbit occurs near perigee. The optimum point must be precalculated and the target orbit parameters must be reduced to

r_T	target radius in the XYZ system
$\dot{r}_T$	target radial rate
V_T	magnitude of the target velocity
Y_T	distance between the plane of the target orbit and the XZ plane

The components of the guidance scheme are algorithms to calculate the time remaining to orbital injection, the required accelerations in the radial and the cross-range direction, and the two command attitude angles.

Time Remaining to Orbital Injection

The powered flight time remaining, T_d, must be estimated to calculate the required accelerations and, subsequently, the command attitude angles. The time predicted is not necessarily required to be very accurate, but must work with the steering law to produce the desired results. This theme will be repeated throughout the paper to obtain a simple formulation. Such an algorithm for remaining time can be obtained by consideration of a well known and very restricted rocket problem.

Let the thrust, F, mass flow rate, $\dot{m}$, and rocket attitude be constant. Neglect gravitational and atmospheric forces. Then the velocity increment over a time, t, can be calculated by

$$V = \int_0^t a\,dt = \int_0^t \frac{F}{m_0 - \dot{m}t}\,dt = V_{ex}\,\ell n\,\frac{\tau}{\tau - t} \qquad (1)$$

where $V_{ex} = \frac{F}{\dot{m}}$ and $\tau = \frac{V_{ex}}{F/m_0}$. Define the velocity yet to be gained in propelled flight as V_d. Then remaining time is given by

$$T_d = \tau\,(1 - e^{-V_d/V_{ex}}) \qquad (2)$$

where V_{ex} is a preset constant and τ is determined from V_{ex} and the measured F/m.

Near perigee, $\dot{r}_T$ is much smaller than V_T so the terminal velocity is approximately horizontal at injection. Further, assume that V_T is locally horizontal throughout flight, this assumption becoming strictly valid only near injection. The radius and radius rate are calculated in the navigator by

$$r = \sqrt{X^2 + Z^2 + Y^2} \qquad (3)$$

and

$$\dot{r} = \frac{\dot{X}X + \dot{Z}Z + \dot{Y}Y}{r} \qquad (4)$$

If V_T, composed of components $\dot{X}_T$ and $\dot{Z}_T$, is locally horizontal, $\dot{r}$ must be zero.

$$\dot{r} = \frac{\dot{X}_T X + \dot{Z}_T Z}{r} = 0 \qquad (5)$$

and

$$V = \frac{X\dot{Z}_T - Z\dot{X}_T}{r} = V_T \qquad (6)$$

since $\dot{Y}_T$ is always zero.

Solving for X_T and Z_T, one obtains

$$\dot{X}_T = -ZV_T/r \qquad (7)$$

$$\dot{Z}_T = XV_T/r. \qquad (8)$$

The gravity losses yet to be accrued could be approximated and added as an additional velocity to be gained; however, sufficient accuracy in estimating T_d can be obtained by ignoring gravity. Therefore, the velocity to be gained is given by

$$\dot{X}_d = \dot{X}_T - \dot{X} = -ZV_T/r - \dot{X} \qquad (9)$$

$$\dot{Z}_d = \dot{Z}_T - \dot{Z} = XV_T/r - \dot{Z} \qquad (10)$$

$$V_d = \sqrt{\dot{X}_d^2 + \dot{Z}_d^2} \qquad (11)$$

This formulation is an approximation covering one rocket stage. The two stage equivalent of equation 2 is easily found to be

$$T_d = T_{d2} + T_{d3}$$

$$= \tau_2 \left\{ 1 - \left[\left(\frac{\tau_3}{\tau_3 - T_{d3}} \right)^{V_{ex3}/V_{ex2}} \right] e^{-V_d/V_{ex2}} \right\} + T_{d3}. \qquad (12)$$

T_{d3} and τ_3 are presettings until third stage ignition as third stage values naturally cannot be derived from measurements during second stage flight. Since T_{d3} is small compared to τ_3 for Saturn class vehicles entering low earth orbit, equation 12 can be reduced to

$$T_d = \tau \left(1 - e^{-V_d/V_{ex}} \right) + T_{d3}. \qquad (13)$$

The two upper stages of the Saturn V vehicle have propulsion systems differing mainly in flow rate so that V_{ex2} and V_{ex3} are equal. Measurements determine $\tau = \frac{V_{ex}}{F/m}$ throughout flight, and T_{d3} is made zero when third stage flight is entered.

The approximation of the remaining time is given by successive substitution into equations 9, 10, 11, and 13. Although these approximations are crude, the performance of the remaining time algorithm is quite accurate. Figure 1 is a plot of percent error in predicted time versus actual remaining time along one of the trajectories used later to determine accuracy and fuel performance. The relatively large error in early second stage flight is due to ignoring the programed step change in fuel flow rate of the second stage. The remaining time algorithm could be designed on a three stage basis to take

the expected variation into account, but guidance scheme performance would not be significantly improved.

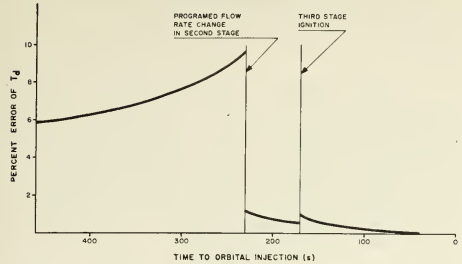


FIGURE 1. ERROR OF REMAINING TIME ALGORITHM ALONG TWO UPPER STAGES OF STANDARD SATURN V TRAJECTORY.

Required Acceleration in the Radial Direction

Guidance of rocket flight is basically a position and velocity end-point control problem. An accurate description of the radial motion as a function of time requires many terms of an infinite series when adhering to the optimum pitch profile. Ignoring the extremum condition, consider the minimum number of terms in a time series for r such that the end-point problem is controllable. The terminal and initial values of r and $\dot{r}$ are determined by the target orbit and first stage performance. The second and higher derivatives of both the initial and terminal values of r are available (unconstrained). The initial conditions of the second and third time derivatives of r are sufficient to control the values of the terminal radius and radius rate. Then

$$\dddot{r} = c \quad (c = \text{constant}) \tag{14}$$

$$\ddot{r} = ct + \ddot{r}_0 \tag{15}$$

$$\dot{r} = \frac{1}{2} ct^2 + \ddot{r}_0 t + \dot{r}_0 \tag{16}$$

$$r = \frac{1}{6} ct^3 + \frac{1}{2} \ddot{r}_0 t^2 + \dot{r}_0 t + r_0 \tag{17}$$

where the subscript 0 means initial. Eliminating c between equations 16 and 17 and substituting $t = T_d$, $r = r_T$, $\dot{r} = \dot{r}_T$ yields

$$\frac{1}{2}\ddot{r}_0 T_d^2 + \dot{r}_0 T_d + r_0 - r_T = \frac{1}{3}\ddot{r}_0 T_d^2 + \frac{1}{3}\dot{r}_0 T - \frac{1}{3}\dot{r}_T T_d. \tag{18}$$

Finally, the required radial acceleration is

$$\ddot{r}_0 = \frac{6(r_T - r_0)}{T_d^2} - \frac{2\dot{r}_T + 4\dot{r}_0}{T_d} \tag{19}$$

Equation 19, when repeatedly solved for $\ddot{r}_0$, prescribes the radial acceleration which is linear in time and which forces $r = r_T$ and $\dot{r} = \dot{r}_T$ when T_d is zero. Equation 19 becomes the steering law for the simplified explicit guidance scheme by dropping the subscript 0 and considering $\ddot{r}$ a command variable:

$$\ddot{r}_c = \frac{6(r_T - r)}{T_d^2} - \frac{2\dot{r}_T + 4\dot{r}}{T_d} \tag{20}$$

Required Acceleration in the Cross-Range Direction

The cross-range problem can be considered identical to the radial control problem. By direct substitution of Y for r in equation 20, one obtains the required cross-range acceleration:

$$\ddot{Y}_c = \frac{6(Y_T - Y)}{T_d^2} - \frac{4\dot{Y}}{T_d}. \tag{21}$$

$\dot{Y}_T$ is zero since the plane of the orbit was assumed to be parallel to the X, Z plane.

Command Attitude Control Law

The vehicle command angles, χ_p and χ_y, are determined from the required accelerations using an integral control law. An integral control law is chosen because of the simple form of the equations and because the command angles are functions of smooth quantities. An added advantage of this method is that the implementation tends to null the effects of misalignment and bias in the attitude control system.

Assume the pitch command angle is given by

$$\chi_p = -k\left[\left(\int_0^T \ddot{r}_c\, dt\right) - \dot{r} + \dot{r}_{co}\right] \tag{22}$$

where k is a constant throughout flight; k should have the largest possible value consistent with an adequate stability margin in the vehicle attitude control system. The $\ddot{r}_c$ integration may be carried out by a simple numerical scheme because integration errors will be corrected by the outer loop (equation 20). Trapezoidal and rectangular integration have both been used successfully. The control law for the yaw command angle may be obtained from equation 22 by substitution:

$$\chi_y = k\left[\left(\int_0^T \ddot{Y}_c\, dt\right) - \dot{Y} + \dot{Y}_{co}\right] \tag{23}$$

The change in sign is required so that equation 23 produces stable motion.

The initial conditions for the two numerical integrators are chosen so that no discontinuity exists in χ_p and χ_y at the beginning of guided flight in the nominal case. The shortest attainable time constant associated with the control law is in the order of ten seconds, resulting in $\dot{r}$ and $\dot{Y}$ following the numerical integrals very well.

Simulation Results

Both the Saturn V and the uprated Saturn I have been examined in simulation studies using the simplified explicit guidance scheme; however, only the Saturn V results are shown. The target orbit was a low circular orbit whose plane contained the launch site. The target parameters were

$$V_T = 7796 \text{ m/s}$$

$$\dot{r}_T = 0$$

$$r_T = 6{,}557{,}166 \text{ m}$$

$$Y_T = 0.$$

Other constants were

$$V_{ex} = 4150 \text{ N s/kg (for both guided stages)}$$

$$k = 0.02 \text{ rad/m/s}$$

$$T_{d3} = 124 \text{ s.}$$

The guidance computations were sampled at two second intervals, and $\ddot{r}_c$ and $\ddot{Y}_c$ were set to zero when T_d reached three seconds. Rectangular integration was used to integrate $\ddot{r}_c$ and $\ddot{Y}_c$. A perfect (errorless) navigator was assumed. Guidance started shortly after second stage ignition; command attitudes were functions of time during first stage flight. The upper stage propulsion systems were represented by constant flow rates and thrusts over prolonged periods of time to simplify the simulation. Thrust termination for the third stage was based on velocity magnitude and V_T only; that is, T_d was not used for engine shutdown.

Nominal data for all stages produced a trajectory which was smoothly controlled and which intersected the target conditions with negligible error. However, the weight delivered to orbit was one-third percent less than the maximum obtainable when adhering to the optimum pitch profile. Further examination showed that the third stage steering was very near optimum; therefore, the performance was being lost almost entirely during second stage flight. Because the steering function for rocket problems of this type is a "shallow" optimum, a function of time was added to force equation 20 to predict the optimum $\ddot{r}_c$ in the nominal data case. Since a minimum fuel trajectory is always required before targeting any guidance scheme, the function of time added to equation 20 can be calculated conveniently along with the optimum trajectory. It is only necessary to execute the remaining time algorithm and equation 20 in the optimization program so that the difference between $\ddot{r}_c$ and the optimum $\ddot{r}$ may be tabulated. Equation 20 should be modified to

$$\ddot{r}_c = \frac{6(r_T - r)}{T_d^2} - \frac{2\dot{r}_T + 4\dot{r}}{T_d} + f_0 + f_1 T + f_2 T^2 \qquad (24)$$

where T is the time from second stage ignition and f_0, f_1, and f_2 are chosen to form a parabolic fit to the error discussed. At third stage ignition, f_0, f_1, and f_2 are set to zero.

Table I contains a tabulation of simulated trajectory results with variation in specific impulse and flow rate in each of the three rocket stages. The variations are taken individually in the upper stages but are combined on a worst case basis along with winds in the first stage. Each perturbation is approximately two standard deviations. Since the loss in payload when using equation 24 is so slight, the simplified scheme can be considered fuel optimum for all practical purposes.

Table I. Accuracy and Performance Summary of the Simplified Explicit Guidance Scheme Applied to Circular Orbit Injection with the Saturn V.

Condition	Injection Radius (m)	Injection Radial Rate (m/s)	Decrease in Weight to Orbit Compared to Optimum (percent)
Standard	6,557,166	0.56	0.05
Second stage I_{sp} low	6,557,166	0.53	0.05
Second stage $\dot{m}$ high	6,557,166	0.52	0.02
Third stage I_{sp} low	6,557,166	0.59	0.05
Third stage $\dot{m}$ high	6,557,166	0.55	0.05
Worst case RSS combination of first stage $\dot{m}$, I_{sp}, and wind variations	6,557,166	0.56	0.07

Conclusions

The complete guidance algorithm, Table II, is given by successive substitution into equations 9, 10, 11, 13, 24, 21, 22, and 23. Note the absence of iteration in these equations. Furthermore, there are no memory elements of any kind except the $\ddot{r}_c$ and $\ddot{Y}_c$ integrators. This fact makes linearization for control system stability studies fairly straightforward. The scheme has excellent accuracy and fuel economy properties as previously shown in Table I. Finally, the simple form of the equations and the small number of vehicle-dependent constants make the simplified explicit guidance scheme attractive in comparison to the older methods.

Table II. Summary of Guidance Equations in Proper Order for Execution.

$$\dot{X}_d = -ZV_T/r - \dot{X} \quad (9)$$

$$\dot{Z}_d = XV_T/r - \dot{Z} \quad (10)$$

$$V_d = \sqrt{\dot{X}_d^2 + \dot{Z}_d^2} \quad (11)$$

$$T_d = \tau\left(1 - e^{-V_d/V_{ex}}\right) + T_{d3} \quad \text{with } \tau = \frac{V_{ex}}{F/m} \quad (13)$$

$$\ddot{r}_c = \frac{6(r_T - r)}{T_d^2} - \frac{2\dot{r}_T + 4\dot{r}}{T_d} + f_0 + f_1 T + f_2 T^2 \quad (24)$$

$$\ddot{Y}_c = \frac{6(Y_T - Y)}{T_d^2} - \frac{4\dot{Y}}{T_d} \quad (21)$$

$$\chi_p = -k\left[\left(\int_0^T \ddot{r}_c dt\right) - \dot{r} + \dot{r}_{co}\right] \quad (22)$$

$$\chi_y = k\left[\left(\int_0^T \ddot{Y}_c dt\right) - \dot{Y} + \dot{Y}_{co}\right] \quad (23)$$

Bibliography

1. MacPherson, D., An Explicit Solution to the Powered Flight Dynamics of a Rocket Vehicle, Aerospace Corporation TDR-169(3126) TN-2, October 31, 1962 (Contract AF 04(695)-169).

2. Chandler, D. C. and Smith, I. E., Development of the Iterative Guidance Mode with its Application to Various Vehicles and Missions, AIAA/JACC Guidance and Control Conference, August 15, 1966.

3. Lawden, D. F., Optimal Rocket Trajectories, Jet Propulsion, December 1957.

4. Breakwell, J. V., The Optimization of Trajectories, Journal of the Society of Industrial and Applied Math., Vol. 7, 1959.

-- NOTES --

No. 67-622

A PERFORMANCE STUDY OF THE LUNAR MODULE (LM) LANDING RADAR SYSTEM

by

J. EICHLER
Grumman Aircraft Engineering Corporation
Bethpage, New York

AIAA Paper
No. 67-622

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA / AUGUST 14-16, 1967

2.03, 8.08, 10.06

A PERFORMANCE STUDY OF THE LUNAR MODULE (LM) LANDING RADAR SYSTEM

J. Eichler
Grumman Aircraft Engineering Corporation
Bethpage, New York

Abstract

The navigation-guidance system of the Lunar Module (LM) spacecraft includes a Doppler radar system used in the lunar landing phase. The radar supplies three velocity component measurements and one range measurement. These measurements are periodically sampled, smoothed, and statistically mixed with the inertial-measurement-unit data in the on-board computer. The resulting navigation system estimates of position and velocity are used by the guidance law equations to calculate thrust and attitude commands. There was a need to verify radar system performance and evaluate its effects on the navigation system, control system, trajectory, and guidance law. A detailed functional math model of the landing radar system was constructed and integrated with associated LM system math models. This resulted in a simulation which was "flown" over many varied trajectories and the results analyzed. The total study is reported in this paper including a description of the systems, a formulation of the exact problem, a description of the method of solution and the study plan, and an analysis of the results. General guidelines are given on problem formulation, math modeling, study plan design, and adaptation to design changes.

I. Introduction

Within the primary guidance navigation and control system (PGNCS) of the Lunar Module (LM) spacecraft, there is a Doppler radar system used in the powered descent phase of the lunar landing mission. The landing radar system (LR) operates from an altitude of 40,000 feet to touchdown and a velocity range of more than 3500 fps. A study of the dynamic operation of the LR system, required to ascertain and evaluate its performance, alone and in conjunction with other parts of the PGNCS system, is described in Section II of the paper.

A set of guidelines is presented in Section III for application to studies of this type.

II. Radar Performance Study

The Lunar Module (LM) primary guidance navigation and control system (PGNCS) has been described in detail in a number of reports and papers (Refs. 2, 4, 9). A brief summary is given here, primarily regarding the functional operation of the PGNCS during the powered descent phase of the lunar landing mission.

A. Brief Description of the LM PGNCS

The powered descent portion of the LM lunar landing mission commences at an altitude of 50,000 feet above the lunar surface and at a central angle of 15 degrees to the nominal landing site. An explicit guidance law brings the vehicle through two guidance subphases to a hover point above the landing site. In the first guidance subphase, the vehicle's speed is reduced from 5500 to 750 fps while the altitude is reduced from 50,000 to 10,000 feet. During the second guidance subphase, the altitude and velocity are reduced further, while also maintaining visibility of the landing site.

The altitude at the end of the second guidance subphase has varied during the course of the study from 100 to 700 feet. The velocity is reduced to a hover condi-

This work was performed under Contract NAS 9-1100 with the Manned Spacecraft Center of the National Aeronautics and Space Administration.

The author gratefully acknowledges the efforts of the many people who contributed to the formulation of the math model, the operation of the simulations and the analysis of the results. Special thanks are given to Bob Harrington of Ryan Electronics for his work on the math modeling of the landing radar, and to Stan Birnbaum, Richard Gardos, John Steinhoff, Sal Sperazza, and Marty Steigman for their efforts.

tion, zero velocity relative to the surface.

The primary sensors of the navigation system are accelerometers mounted on a gyro-stabilized inertial platform. The accelerometer data are sampled and used in the on-board computer within the equations of motion to ascertain a best estimate of position and velocity. These estimates are utilized within the equations of the guidance law to compute new thrust and attitude commands. The guidance equations and the equations of motion are solved at the rate of once every two seconds. A digital autopilot is used to maintain the rotational stability of the craft and to attain the required thrust positioning as commanded by the guidance law. Complementing the above system is a continuous wave Doppler radar which is turned on at 40,000 ft altitude. Three radar beams measure velocity and the fourth measures range (altitude). They are mounted in a lambda configuration (see Fig. 4). The radar antenna is attached below the ascent stage of the craft and has two angular positions available. The angle switching logic and other functional aspects of the radar operation are explained in sections D and E. Between 40,000 and 25,000 ft, radar data are available for astronaut monitoring but are not used in the auto - matic system. From 25,000 ft and down, the radar data are used by the PGNCS in the manner described below.

The functional flow of the PGNCS, including the LR, is shown in Fig. 5. The flow of data is from the LR to the interface block, and then to the navigation equations of motion, the guidance law, and finally the vehicle block, the outputs of which are sensed by the landing radar. The data sampling schedule is:

Above 25,000 ft altitude - no sampling

Between 25,000 and 15,000 ft altitude - every 10 seconds, 2 samples are taken of range at 80 millisecond intervals

Below 15,000 ft altitude - every 10 seconds a complete sampling of range and three velocity components takes place

The sampling sequence consists of range followed by the X, Y, and Z velocity components at 2-second intervals. There is a 4-second interval between the Z velocity update and the next range update resulting in a total sampling period of 10 seconds. The sampling of each velocity variable consists of 5 samples taken at 80 millisecond intervals.† This schedule is represented in Fig. 1.

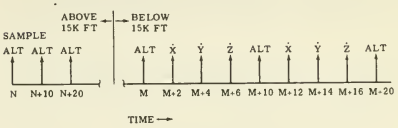


FIGURE 1 - RADAR SAMPLING SCHEDULE

Subsequently, this schedule was altered below 15K ft to the following:

FIGURE 2 - RADAR SAMPLING SCHEDULE

Altitude sampling is inhibited 20 seconds before the end of each guidance subphase. This inhibiting is required to maintain proper functioning of the explicit guidance law, because if a position correction is received by the guidance law with little time-to-go, it would create an unmanageable situation.

The radar data are taken in sets of five samples for velocity (and two samples for altitude) which are averaged and then weighted with the present navigation estimates as follows:

†The description of the LM landing radar system reflects the configuration at the time the study was conducted and should not be construed to represent the present configuration.

Range - a linear weighting factor is applied with a value of zero at 25,000 ft. altitude and a value of 1 from 5,000 ft. altitude down.

Velocity - a linear weighting factor is applied with a value of zero at 15,000 ft. and a value of one at zero altitude.

These linear weighting factors were later altered to the following, also linear, weighting factors:

$$W_{vxa} = 0.40\left(1 - \frac{V_{TOT}}{1550}\right)$$

$= \dot{X}$ component weighting factor

$$W_{vya} = 0.70\left(1 - \frac{V_{TOT}}{1550}\right)$$

$= \dot{Y}$ component weighting factor

$$W_{vza} = 0.70\left(1 - \frac{V_{TOT}}{1550}\right)$$

$= \dot{Z}$ component weighting factor

$$W_A = 0.55\left(1 - \frac{h}{25000}\right)$$

= altitude weighting factor

where V_{TOT} = the estimated total velocity of the craft and h = the estimated altitude of the craft.

In both cases, the updated navigation estimate is obtained using the following equation:

$$\dot{X}_w = \dot{X} + W_v(\dot{X}_{LR} - \dot{X})$$

where

$\dot{X}_w$ = the new weighted estimate of $\dot{X}$

$\dot{X}$ = the navigation estimate of $\dot{X}$

$\dot{X}_{LR}$ = the sampled, averaged value of $\dot{X}$ derived from the LR

W_{vx} = the weighting factor for $\dot{X}$

The same equation is cycled for $\dot{Y}$ and $\dot{Z}$, changing the appropriate subscripts. In the case of altitude, W_v is replaced by W_A, and $\dot{X}$ is replaced by h, altitude. The updated estimates are then used by the guidance law to calculate new thrust and attitude commands. The estimates also serve as a new set of initial conditions for the equations of motion.

There are two phenomena affecting radar operation which require a brief description. They are the lunar reflectivity and the zero Doppler effect. In analytically determining the return energy reflected from the lunar surface and received by the radar, an assumption is made regarding the reflectivity (the fraction of reflected energy) returned by the lunar surface. Since the Surveyor spacecraft made the first partial measurements of lunar reflectivity, the value as a function of surface incident angle was a matter of some conjecture (see Ref. 10). Some estimates of this critical parameter are shown in Fig. 7.

The zero Doppler phenomenon is the loss of signal energy when the radar beam is orthogonal to the velocity vector of the craft. A close approach to orthogonality is sufficient to reduce the signal enough to be indiscernible from the noise, i.e., reduce the signal to noise ratio (S/N) below detection level, and cause loss of track. The deviation from orthogonality sufficient to cause zero Doppler track loss varies from a ±15° band about the velocity vector normal at high altitudes to a few degrees at low altitudes. It is a function of velocity, incidence angle, signal power, radar mode, and bandwidth.

B. Performance Requirements and Criteria

There are two types of performance requirements and criteria: those specified for the radar alone, and mission requirements. The first type is found in Ref. 7, parts of which are repeated below. The accuracy specification is given in Table 1.

Additional performance requirements are: radar design limits specified in Ref. 7, must not be exceeded; altimeter data must be available below 25K ft altitude, and velocity data should be available from the start of guidance subphase two; zero Doppler regions must not be excessive.

Within the mission requirements, the main criteria is To Land Within The Touchdown Velocity Envelope (see Fig. 6). Mission requirements also include the following: ΔV budget compliance; visibility of the landing site compliance during guidance subphase two; compliance with the state vector aim point at the end of guidance subphase two (see Part A for a discussion of the guidance subphases); and landing within a 1/2 nautical mile CEP of the landing site.

Altitude (ft)	25000-2000	2000-200	200-5
V_{xa}	*1.5% or 1.5 fps	1.5% or 1.5 fps	1.5% or 1.5 fps
V_{ya}	2.0% or 2.0 fps	3.5% or 3.0 fps	2.0% or 1.5 fps
V_{za}	2.0% or 2.0 fps	3.0% or 3.0 fps	2.0% or 1.5 fps
Altitude (ft)	25000-2000	2000-10	
Range	1.4% + 15 ft	1.4% + 5 ft	

Table 1 LANDING RADAR ACCURACY SPECIFICATION

*Note: Percentages are of total velocity for the velocity specification and of the true altitude for the altitude specification.

C. Statement of the Problem

With the knowledge of the functioning of the systems within the PGNCS, including the landing radar system, how may we study and eventually answer two questions, Does the radar system operation comply with radar system requirements? and Does the radar system operation, in conjunction with the other systems within the PGNCS, comply with mission requirements?

The PGNCS system has been briefly described and the radar and mission requirements given. This, together with the two preceding questions, state the problem. However, other questions arise subsidiary to these questions.

What effects are produced, if any, by the coupling of errors developed in the radar system through the guidance and control system? Since the LR accuracy is sensitive to attitude, does the effect of radar measurement errors cause the control system to command attitude changes which increase the radar measurement errors, thus creating a positive feedback situation?

What effects do LR errors have on meeting the state vector aimpoint at the end of subphase two? This is to be considered from the points of view of the navigation estimate miss and true miss.

How often and how long is radar data not available because of zero Doppler?

What are the effects of different lunar terrains; of different lunar reflectivity; of different initial condition errors (IC); of different gyro drifts, accelerometer biases, and other sensor errors; of different errors in the thrust levels achieved by the descent engine; and finally, of different maneuvers by the astronauts to select an alternate landing site?

A discussion of the method applied to answer these questions follows.

D. The Method of Solution

There are a number of theoretical ways to approach the problem stated in Part C. These are:

Closed Form Analytic Methods - The seeking of closed form solutions to determine the effects of different errors in the radar and other PGNCS on the mission criteria is not feasible in this situation. The complexity and nonlinearity of the systems involved preclude use of closed form solutions without making simplifying assumptions that contradict the accuracy of solution required by the problem.

Flight Test - A series of flight tests were conducted to obtain measurements of the radar system operation. These were done on helicopters and other aircraft and are of limited use in answering the problems posed in Part C "Statement of the Problem." The limitation arises from the difference of reflectivity of the earth, the incomplete configuration of the PGNCS system flown, and the inability in general to match lunar environment in the trajectory, atmosphere, gravity field, etc.

Static Test and Analysis - Testing of all the components of the PGNCS has and is being done, and this furnishes necessary data for the following method. It cannot, however, solve the problems of dynamic performance. The analytic solution, point by point, of the landing radar measurement

errors (conducted at Ryan, see Refs. 1, 3) is useful but lacks the dynamic characteristic, that is, the feedback effects, required to answer the problem.

Dynamic Analysis Employing a Simulation - A simulation which represents the LM spacecraft and includes math models of all the systems directly associated with the problem, affords the possibility of "flying" closed-loop-system trajectories. This yields the dynamic response of all systems, and is limited only by the accuracy and detail of the math models and computers.

The last method, Dynamic Analysis Employing a Simulation - was selected for studying and solving the system performance problem. There was in operation at Grumman a six degree of freedom simulation of the LM guidance and control systems (see Ref. 8). It contained functional math models of the entire PGNCS with a simplified model of the radar system. Since construction of an adequate math model of the landing radar system and integration into the simulation would make it adequate for studying the performance problem, a landing radar math model was constructed. This model is discussed in the following two subsections.

1. Rationale and Assumptions. The basic ground rule which was applied in the development of this math model (MM) is that the LR system is described functionally without concern for the effect of each electronic or mechanical component. The LR system was considered a black box with input and output and a series of equations describing the internal functional operations. This is in contradistinction to a part-by-part MM which would describe the function and effect of each component. The math model did not include the LR display devices in the LM cabin.

The computer restrictions on computation time and memory storage allowed the retention of primary error sources only and required that they be represented in a simplified mathematical form.

The primary errors are classified in three categories. Bias errors, including terrain bias, dynamic lag, Doppler compensation and the preamp slope effects, are constant offsets varying from point to point in the trajectory, and are functions of the state vector. Installation errors are derived from the uncertainty of the final installation of the radar on the LM and of the uncertainty of the alignment of the descent stage on the ascent stage. These are considered to be constant for the duration of the run, that is, for powered descent from 40 K ft to touchdown. Fluctuating errors, spread spectrum, VCO drift, modulation rate and quantization effects, are randomly varying errors and their 1σ value is a function of the state vector at each point. The assumption of linearity was applied in combining these three types of error sources. It was also assumed that the random error sources have a gaussian distribution and are uncorrelated. The major error sources are discussed individually below:

Installation errors - this error source includes the effect of uncertainties in the mounting of the radar on the LM, the orientation of the beam directions on the antenna plate, and maximum vehicle distortion resulting from vibration and temperature changes.

Terrain bias errors - this error source is the shifting of the return signal peak caused by impingement of the finite width beam on a surface of varying reflectivity.

Preamp slope errors - this error source is due to the shifting of the return signal spectrum peak by uneven gain as caused by the slope of the preamp characteristic curve.

Spread spectrum error - this error is caused by the uncertainty in the bandwidth of the return signal. Spreading of the bandwidth is a function of the bandwidth itself and the post detection smoothing time, and results in an uncertainty in knowledge of the center Doppler frequency.

VCO errors - the drift of the voltage controlled oscillator (VCO) is the source of this error.

Dynamic lag errors - the dynamic lag of the landing radar is known in part and will be compensated in the LM guidance computer (LGC). The model includes only the uncompensated part of the dynamic lag error.

Doppler compensation errors - in the landing radar, the frequency shift must be subtracted due to the velocity component along the range beam from the total range beam frequency. To accomplish this, the Doppler frequencies of the two beams which straddle the range beams are averaged. The electronics introduce a fixed scale factor error, which results in a range error.

Modulation rate uncertainty errors - the source of this error is the uncertainty in the modulation rate of the range beam frequency.

Quantization errors - prior to leaving the landing radar system, the detected and filtered Doppler and range frequencies are digitized in an analog to digital converter. There is a resulting uncertainty caused by the finite resolution of the converter.

A simplifying assumption made in the derivation of the model is that the return frequency has a gaussian shape centered about the mean frequency. This is a standard assumption made in analyses of this type.

With regard to the completeness of the model, the following errors were not included, although they may be primary type errors:

Plume effects - this error source was neglected because there is no accurate model to describe this phenomenon at the present time.

Reflections from the landing gear and engine nozzle extension - although an exact model of these error sources was not made for the same reason as plume effects, a gross noise estimate was made to cover this error source. Later in the study, Tables 2 and 3, measured data were available concerning these error sources and were included.

Temperature and pressure variations - the maximum effect on misalignment due to these sources was included, but the effects due to their rate of change during the mission have been neglected. The effect of their rate of change on the LR system is assumed to be negligible.

Sidelobes - this effect was omitted from the math model because it is being studied separately. The results of this study will indicate whether this effect should be subsequently added to the math model.

Track requirements hysteresis - a different S/N ratio is required for maintaining track. In this model, because of computation time and computer memory restrictions, only one table of S/N requirements was included, namely, the table for maintaining track. This assumption was removed in the studies detailed in Tables 2 and 3.

Antenna-switching time delay - this effect was not included directly due to computer time and memory restrictions. However, its effect in causing loss of track was considered.

2. Description of the Equations. A simplified block diagram of the landing radar MM is shown in Fig. 5. The math model consists of four major parts which are briefly described in the following paragraphs.

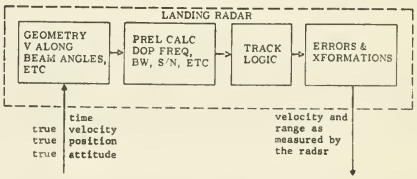


FIGURE 3- LANDING RADAR MATH MODEL FLOW CHART

The input to the math model includes true velocity, true position, and true attitude. In the geometry block, the following variables are calculated: true velocity along the radar beams; true direction cosines between the LM body axes and the radar beams; the angles between the radar beams and the local vertical; the range along the beams; the angles between the radar beams and the velocity vector; and the angle between the velocity vector and the local vertical.

In the second part of the radar math model, entitled "Preliminary Calculations," the following variables are computed: Doppler frequencies along the radar beam; sensed range frequency; bandwidths; signal power, noise power, and signal to noise ratios (S/N); open loop tracker gains; and smoothing time.

In the third section of the model entitled "Track Logic," a determination is made of whether each radar beam is in "track." This determination follows the corresponding functioning of the electronics within the radar, which compares the signal to noise ratio for each tracker beam, decides whether it is sufficient to

accept the signal as accurate, and accordingly is "on track" or has "lost track." In the latter case, a search routine of the operating frequency region is conducted, seeking the lost signal. When the signal is found, the radar "locks on." The search pattern time is 6 seconds for one complete sweep and in the simulation it was calculated that track was lost for a minimum of 6 seconds, when track loss occurred. After the minimum 6-second period, track was calculated to have been regained as soon as sufficient S/N ratio occurred. The matter is further complicated by there being two modes of operation requiring different S/N ratios. The discussion of this is found in Ref. 7.

In the final section, a calculation is made of the following errors: spread spectrum error; installation errors; terrain bias error; dynamic lag error; preamp slope error; VCO drift error; modulation rate error; and Doppler compensation error.

Finally, the error components are summed along each radar beam and are added to the true values. The measured velocity and range values are then transformed into coordinates fixed in the LM vehicle. This transformation is not included within the math model, but is performed in the interface block.

The integration routine utilized throughout the LR model was a single point slope algorithm used in an integration interval which allowed 10 equal divisions of the area to be integrated. A random number subroutine called BOXNO, and the usual trigonometric square root and logarithmic subroutines of the 7094, were called as required in the program. These are on the Grumman Monitor Library Tape for the IBM 7094 computer.

The math models which complement the landing radar model, to comprise a total simulation, include the following systems:

The guidance law

The navigation system; sensors and equations of motion

The radar-navigation interface

The control system; the digital autopilot

The propulsion system including the main engine (descent engine) and the reaction control system jets

The vehicle dynamics: including the rotational equations of motion; and the translational equations of motion; ΔV computation; visibility of the landing site computation; and gimbal lock equations.

The equations of the landing radar math model are in Ref. 6. The above list refers to the six degree of freedom guidance and control verification study simulation which was mechanized on a hybrid computer complex. For later studies, as discussed in Part E, "The Study Plan," the same models were used except that a simplified control system was used in place of the digital autopilot. This simplified system was a "perfect" control system with rate limiting. In addition, the rotational equations of motion were simplified to assume perfect response with rate limiting. The second simulation was mechanized solely on digital computers. A flow chart of the simulation is shown in Fig. 8.

E. The Study Plan

The decision was made to use simulations of the PGNCS to study and answer the questions posed in Part C, "The Statement of the Problem." Two different simulations were chosen. The first to be used was the guidance and control verification simulation which was augmented to contain a very complete landing radar math model as described in Part D "Method of Solution." Because this simulation had a complete representation of the control system, the digital autopilot, and because it had a sophisticated analog method for solving the rotational equations of motion, it was used to study the coupling of landing radar measurement errors with the control system. When this problem was solved (see Part F "Results and Conclusions") the second, simplified, all-digital simulation was employed. This simulation is described in Part D "The Method of Solution." Note that only the automatic mode of the LR system was studied.

The matrix of runs made on the first simulation is tabulated in Table 2, and titled "Matrix of Runs I." Run 1 was a nominal, error-free trajectory. Run set 2 gave the effect of 1σ installation errors. Run set 3 considered the effects of fluctuating errors and run set 4 studied bias errors. In run set 5, the additional effect of artifically forcing the trajectory to have a positive altitude error at radar turn-on was studied, showing the effect of each source of radar error separately. Run set 6 studied the same situation with a low off-nominal altitude error. Runs 7 and 8 were used to study the possibility of improved radar operations by changing to a new antenna mounting angle.

Runs	Description	Error-Free	LR ERRORS INCLUDED: Installation Errors Only	Bias Errors Only	Fluctuating Errors Only
1	Nominal Trajectory	X			
2	Nominal Trajectory with 1σ installation errors		X		
3	20 runs of Nominal Trajectory with randomly varying fluctuating errors only				X
4	Nominal Trajectory with bias errors only			X	
5	High Off-Nominal Trajectory; repeat steps 1, 2, 4 and run fluctuating errors one run only at 1σ value	X	X	X	X
6	Low Off-Nominal Trajectory; repeat as in 5 above	X	X	X	X
7	New Antenna Angle Nominal Trajectory, error-free	X			
8	New Antenna Angle Low Off-Nominal Trajectory, error-free	X			

Table 2 MATRIX OF RUNS I

Runs	Description	CONDITIONS: Antenna Angles	Reflectivity	Pitch Misalignment
1 1.1.4	A Family of Trajectories with nominal reflectivity and varying pitch misalignment	40, 0 degrees	nominal	0
1.1.1				-5 mr
1.1.3				-2.5 mr
1.1.5				+2 mr
1.1.6				+4 mr
1.1.8				+16 mr
2 2.1.1-2.1.8	The above set repeated with increased reflectivity assumed	40, 0 degrees	increased by 5db	
3 1.1.11-1.1.18	Set 1 repeated with antenna angles of 50.7° and -20.4°	50.7°, -20.4°	nominal	
4 2.1.11-2.1.18	Set 3 repeated with increased reflectivity assumed	50.7°, -20.4°	increased by 5db	
5 3.1.4	Various antenna angle strategies	0°, 0°	nominal	0
4.1.4		0°, 0°	increased by	0
5.1.4		15°, 15°	5db nominal	0

Table 3 MATRIX OF RUNS II

The matrix of runs made with the second simulation is tabulated in Table 3 and Table 4, and is titled "Matrix of Runs II," and "III," respectively.

"Matrix of Runs II" was designed to determine the performance of the PGNCS with the antenna mount angles changed from 50.7 and -20.4 degrees to 40 and 0 degrees, respectively. The angles refer to guidance subphase variation, there being a mount angle change made at the change in guidance subphase. The first design called for 50.7 degrees for guidance subphases one and two and -20.4 degrees for the descent from hover. The second design called for 40 degrees for guidance subphase one, and 0 degrees from there to touchdown. The need was to evaluate the change in radar and PGNCS performance resulting from this antenna angle design change. In addition, there was a need to study the variation of operation resulting from assuming a 5db increased reflectivity of the lunar surface. An indication that the lunar reflectivity may be higher was given by Surveyor data. The run Matrix II also reflects changes in the math model resulting from test results data which showed a higher noise power than had been previously estimated. Increased noise also increased the zero Doppler dropout occurrences. The matrix evaluated these changes by "flying" a nominal trajectory and a family of high and low off-nominal trajectories induced by pitch misalignments of the platform. This method of inducing high and low trajectories was selected because of its simplicity and because it was found in previous studies that pitch variation is the major cause of zero Doppler occurrence. See Part F "Results and Conclusions" for details. In run set 5 of Matrix II, several constant angle antenna strategies were studied while flying nominal trajectories only.

The final study of antenna angles was made in run set 3 Matrix III shown on Table 4. In this set a series of different angles were considered for subphase 1, since it was discovered in run set II that 40 degrees did not yield optimum performance. In run set 1 of Matrix III, redesignation to new landing sites was studied by selecting three points on a circular footprint of available landing sites, namely, the right edge, left edge, and downrange edge. The nominal is the lower edge. These are shown in Fig. 9. In Section D of Series 1, sensor errors were introduced.

1.1 included gyro errors in each axis of drift = ±.32 deg/hr, misalignment = ±3mr.

1.2 included accelerometer errors in each axis of bias =.75 cm/sec^2 scale factor = 750 pp million.

1.3 included thrust variations of ±3 percent at engine start to ±4 percent at throttle down time.

1.4.1 and 1.4.2 were selected initial condition errors in altitude, cross range, downrange and their velocities (Ref. 2, cases #212 and #1191).

1.5 included a selected set of combined errors. Under series 2, two different terrain models were flown. These are reproduced in Fig. 10.

F. Results and Conclusions

The results refer to the matrix of runs I, II, and III, and are discussed accordingly.

1. Results of Matrix I. To answer part of the question, "Does the radar system operation comply with the radar system requirements?", it was necessary to reduce the data achieved through runs I, and compare them with the specified accuracy.

The three sources of error, fluctuating, installation, and bias were added in the following conservative manner. The root sum square of fluctuating plus installation errors was added to the magnitude of the bias errors, and the sign of the bias error was assigned to the sum. Absolute values were used for computing percent accuracy.

Track loss was computed in all trajectories, including those designated as error-free. The 20 runs with random fluctuation errors were analyzed to obtain the mean and standard deviation and for certain points the chi-square test was applied to see how close the distribution was to a normal one.

Accuracy Curves: In the graphs shown in Figs. 11 through 22, the accuracy results are depicted for nominal, high off-nominal and low off-nominal trajectories. Track loss periods and the accuracy specification are indicated on the graphs.

The accuracy tables indicated that compliance with the specification was achieved on all trajectories with the exception of range in the low off-nominal trajectory. This trajectory is not a normal LM descent trajectory and is representative of a trajectory requiring an abort. It was utilized only for the

Series	Conditions and Purpose
1	Redesignation trajectories w/wo IMU and IC errors and Thrust Variations
	a) Redesignation to right edge of the allowable ellipse of landing sites, nominal condition
	b) Redesignation to left edge of the allowable ellipse of landing sites, nominal condition
	c) Redesignation to downrange edge of the allowable ellipse of landing sites, nominal condition
	d) Repeat series 1a with errors.
	1.1 series gyro errors
	1.2 accelerometer errors
	1.3 thrust variation
	1.4.1 I.C. case #212*
	1.4.2 I.C. case #1191*
	1.5 a select combination of errors
	e) Redesignate downrange with a select set of errors
2	Lunar Terrain Effect Trajectories, using models T-6 and T-8
	a) T-6 Terrain model, run with nominal conditions and select error cases from series 1
	b) T-8 Terrain model, run with nominal conditions and select error cases from series 1
3	Antenna Angle Evaluation, use conditions selected from series 1 and 2 to evaluate a new antenna angle, if required.

*Table 5.6 MIT Report E 1982, Ref. 2.

Table 4 MATRIX OF RUNS III

purpose of showing LR update performance under extreme conditions. It is noted further that these accuracy curves indicate a large margin of accuracy available for most of the trajectory. The closest approach to specification values occurs at about 200 ft, which is the end of subphase two in these trajectories. It is indicated by the relatively small variation in the accuracy for all the trajectories considered, that the feedback of LR errors through the balance of the system is small. If feedback were an important factor, a sharp deviation in accuracy would have been experienced at some point in the off-nominal trajectories.

Sensitivity at hi-gate and lo-gate: LR errors were introduced at the 1σ value and the variations of the navigation estimate from the error-free case was calculated at hi-gate and lo-gate. The navigation errors were calculated separately as resulting from installation, bias, and fluctuating errors, and the data were assembled as indicated above. This process was repeated for trajectory deviations at lo-gate. The term "sensitivity" is used here to indicate the bias $\pm 1\sigma$ variation in navigation (or trajectory) values resulting from bias, fluctuating and installation errors in the radar system. The landing radar fluctuating and installation errors are taken at their 1σ value. The coordinate system is approximately x-local vertical, y-out-of-plane, and z-down range.

There is no sensitivity specification except for the constraint that the landing velocity envelope be achievable. The data for end-of-subphase-one (hi-gate) sensi-

tivity (Fig. 23) show very small errors, less than 30 feet in position coordinates and less than 5 fps in velocity coordinates. The data for end-of-subphase-two (Lo-gate) (Fig. 24) indicate that the landing velocity envelope is achievable. The navigation errors in position are less than 300 ft and the velocity errors less than 1 fps. Therefore, the navigation estimate of velocity is only slightly off the true value and, with subsequent continued updates, even these small errors should be reduced. The position errors are obviously well within the half mile CEP. The trajectory errors at this point are also small and should be partially decreased due to present good navigation data at the end-of-subphase-two and anticipated improved navigation data, during the balance of the descent. It is concluded, therefore, that navigation and trajectory sensitivity to LR errors will allow a successful landing, within the velocity envelope and 1/2 mile CEP.

Track Loss Information: In Fig. 25, the effect is shown of various error sources on track loss. It is apparent that no increase in the frequency of track loss occurred and that the total effect was only to shift the time of commencement of track loss by small amounts. It is, therefore, concluded that the effect of errors on track loss is negligible.

In Fig. 26, the event of track loss is depicted for the nominal trajectory. At a pitch angle of 68°, approximately, the region of zero Doppler is entered. It is left at approximately 50° pitch angle. Track loss, once initiated, continues for a minimum of six seconds in the simulation. The six seconds represent the nominal search pattern time for one complete pass. After six seconds the S/N ratio is tested, and reacquisition occurs as soon as the required criteria are satisfied.

In Fig. 27, the true altitude of the LM is plotted versus time for the nominal trajectory. The arrows indicate pitch angle measured counterclockwise from zero toward the top of the paper. Track loss is indicated by the heavy line. It is noticed that velocity track is available below 26.4 K feet. Range data is available from above 40 K feet.

In Figs. 28 and 29, the true altitude of the LM is plotted versus time for a high off-nominal trajectory and a low off-nominal trajectory, respectively. The same graph format is used here as for the nominal trajectory. In addition, the dotted line is drawn to show the navigation estimate of altitude. The numbers adjacent to the dotted line indicate the number of the altitude update. After very few updates the navigation estimates are so close to the true value that the difference cannot be depicted on the scale of these graphs. This demonstrates graphically the effectiveness of the LR system in correcting navigation data in highly off-nominal conditions. Track loss occurs frequently in both these extreme trajectories, as indicated.

A family of error trajectories is shown in Fig. 30. This family of curves indicates the trend in track loss occurrence as more highly off-nominal trajectories are flown. The errors in pitch misalignment which were utilized to produce these off-nominal trajectories are indicated on the graph. It can be seen that the pitch misalignment must be at least 4σ or -10σ to cause an increase in frequency of track loss above that experienced on the nominal trajectory. The continuity of variation of the frequency of track loss occurrence is also indicated. From this data it can be concluded that increased track loss occurrence due to pitch misalignment alone, above that obtained on the nominal trajectory, is not probable and, therefore, need not be considered a serious cause of error.

S/N Data: In Fig. 31, the S/N data is presented for the nominal trajectory for beams 1, 2, and 3 of the radar. The crosshatched area indicates the boundary where the S/N ratio falls below that required for track. At 26.4 K feet beam 3 emerges from the no-track region and satisfies the requirement for velocity data. Between 30 K to 40 K ft, though not depicted, beams 1 and 2 are in track, thus satisfying the requirement for altimeter track. In the system, however, altimeter data is only used below 25 K feet and velocity data below 15 K feet. The requirements for altimeter data and velocity data are complied with for the nominal mission.

Alternate Antenna Angle Data: For the purpose of studying performance with a different LR antenna angle (43°), runs 7 and 8 (Table 1) were made. At the time of this study, the antenna angle of 43° was being considered in order to reduce certain spurious noise problems caused by reflections off the engine nozzle and landing gear. Subsequently, an angle of 40° was selected. Nevertheless, the data obtained at 43° was useful as a first cut rough indication of the results to be anticipated at 40°. The final selection

is reported under Matrix III.

A comparison of S/N data for the nominal and the new antenna angles for the nominal trajectory is given in Fig. 32. It is seen that with the new antenna angle (43°), beam 3 remains in the track loss area (crosshatched) down to an altitude of 23 K feet compared to 26.4 K feet with the nominal antenna angle. Beams 1 and 2, however, achieve higher S/N ratios at earlier time points, i.e., higher vehicle altitudes.

In Fig. 33, the effect of the 43° antenna angle on track loss and altitude is shown for the low off-nominal trajectory. This graph should be compared with Fig. 29, which depicts the same data for a nominal antenna angle.

The track loss performance for the two angles appears to be similar, and no conclusion can be drawn from the small amount of data available. A determination of the optimum antenna angle is the subject of the study in matrix III.

ΔV Consumption: It was observed that for the nominal trajectory the additional ΔV consumption due to LR errors did not exceed 2.5 fps.

2. Conclusions from Matrix I. - The following conclusions result from this study:

a) The 3σ radar accuracy specification is complied with on all runs, except range on one highly off-nominal trajectory.

b) Track is available at 40 K feet for range and 26.4 K feet for velocity on the nominal trajectory. This is in compliance with the specification requirements.

c) A six-second track loss occurs during the nominal trajectory during pitchover at hi-gate.

d) The navigation accuracy and trajectory dispersions resulting from LR errors are reasonable and can allow a successful touchdown.

e) The LR error feedback through the guidance loop produces negligible system deterioration.

f) The LGC sampling, smoothing and weighting system for the LR data performed satisfactorily for the trajectories studied.

g) The ΔV consumption change due to LR errors is negligible.

3. Results of Matrix II. On Fig. 34 the results are tabulated showing the comparison in performance between the antenna angles 40, 0, and 50.7, -20.4, both with and without increased reflectivity. The integrated noise level was 10 mv for this data, an increase from the 3 mv assumed for Matrix I. An interface design change was made to increase the altitude sampling rate to once every two seconds. Nevertheless, the study was made at the previous rate of once every ten seconds because the change became effective after the study was partly completed. The data clearly indicates the superiority of the 40, 0 degree antenna angles over the 50.7, -20.4 degree angles, both with nominal reflectivity and considering an increase of 5db.

A series of accuracy runs were made and the results indicated good compliance with the accuracy specification. This is shown in Figs. 35 to 38. The change in S/N ratio is shown in Figs. 39 and 40. Figures 41 to 44 show loss of track areas as a function of altitude error. These graphically depict the reaction of the guidance law to various altitude errors at 25 K feet when the radar data is first used, and were produced by pitch misalignment errors. In Fig. 45 is shown the updating effect on the navigation estimate of altitude by successive radar measurements. Finally in Fig. 46 the nominal trajectory is tabulated.

4. Conclusions from Matrix II. Antenna mount angles of 50.7° and -20.4° are not acceptable because of excess zero Doppler track loss. Antenna angles of 40° and 0° are acceptable. They comply with accuracy requirement and enable landing within the touchdown velocity envelope.

An increase of 5db in the lunar reflectivity greatly enhances the radar operation for both antenna angles sets.

5. Results of Matrix III. Run set 3 of Matrix III was a study to determine the optimum antenna angle for guidance subphase one. For Matrix III several new design decisions were incorporated. They are: modification of the preamplifier roll-off; use of new noise information resulting from radar set testing; and changing the interface to the new weighting factors and sampling rates indicated in Section A.

The resulting data is summarized for the final selected angle of 24° and the various conditions considered in run Matrix III (see Ref. 11).

For the nominal trajectory (see Figs. 47 and 48):

Antenna Angle Above High Gate	Altitude Data Acquisition	Velocity Data Acquisition
24°	26,000' to touchdown	15,000' to touchdown
40°	29,000'; drops out between 20,000' and 11,000'; back in from 11,000 to touchdown	11,000' to touchdown

Reference to Figs. 47 and 48 shows that an antenna angle of 40° above high gate results in an obviously poor design situation compared to 24°. The 24° angle yields sufficient Doppler drop out margin while the 40° angle penetrates the drop out zone. The 40° antenna angle results in successful landings for all cases so far considered with certain deficiencies (see Table 5) which are not present for the 24° angle. Altitude errors at high gate (9,252 ft) are reduced from approximately 2,000 ft for the 40° angle to ±300 ft for the 24° angle. There is a loss of landing site visibility below high gate for certain trajectories.

RCS fuel consumption varies less than ±1 pound for all trajectories and both the 24° and 40° antenna angles. For the 40° antenna angle, the altitude data good light will come on at 29,000 feet and then go out for a prolonged period between 20,000 feet and 11,000 feet. If it does not come on at the expected 11,000 ft altitude, valuable diagnostic time will have been lost, not to mention the astronaut concern involved with potential occurrence of this contingency. Prolonged loss of track between 20,000 and 11,000 ft with the 40° angle is a bad practical radar situation, since vulnerability to false lock from internal sources or side lobes is increased in the search mode.

Consider the 40 degree Doppler drop out plot, Fig. 48: altitude data is acquired at 29,000 ft. However, the LGC weighting factor is zero until 25,000 ft, and only rises to 0.1 when drop out occurs at 20,000 ft. In effect, this means that altitude corrections are only partially (approximately 66%) processed by the LGC when drop out occurs. Effectively, then the 40 degree antenna angle yields little or no altitude update data above 11,000 feet for the cases studied. Thereafter, altitude errors again build up until lock is achieved at 11,000 feet. In one example, pitch misalignment of -3 mr, the altitude error was reduced from 900 to 300 ft in the region 25,000 ft to dropout. Thereafter it increased to 2,000 ft during the zero Doppler region traverse. The pitch angle plot for an antenna angle of 40 degrees, Fig. 49, shows the effect of the prolonged lock loss. Pitch angle excursions, which are indicative of the existence of trajectory altitude corrections in process, are virtually absent above high gate in comparison to the case shown for the 24 degrees antenna angle (Fig. 50). The result is that pitch angle excursions, which yield trajectory corrections, are deferred until after high gate, for the 40 degrees antenna angle. This leads to loss of visibility of the landing site in that period for certain cases. Furthermore, the altitude at which high gate pitch over occurs is widely scattered due to the uncorrected altitude error residue from the high altitude lock and further accrued altitude errors due to the drop out period. In comparison, the 24 degree pitch angle plot, Fig. 50, shows that all errors have been corrected by high gate, resulting in pitch over consistently occurring at the correct altitude with relatively uniform pitch profiles prevailing after high gate.

At a meeting with MSC personnel on January 9, 1967, during which the data were presented, a request was made that seven trajectories be rerun with radar data suppressed until a 20,000 ft altitude was reached (as opposed to the present 25,000 ft altitude). The request was based upon the requirement to allow 30 to 60 seconds for the astronauts to evaluate the data before allowing it to enter the LGC. This implies a maximum of 5 K ft after acquisition of altimeter track at 26 K ft resulting in the figure of 20 K ft. The data is summarized in Table 5 in the column headed 20 K. The results show that only a negligible change in PGNCS performance results from this change for the cases considered.

6. Conclusions from Matrix III. The angles of 24° and 0° have been shown to be in compliance with radar performance

TRAJECTORY	ALTITUDE ERROR AT HIGH GATE (9,252 FT)			LANDING SITE VISIBILITY LOSS AFTER HIGH GATE ALTITUDE BAND IN FT/ TIME INTERVAL IN SEC			TOTAL ΔV EXPENDED IN DESCENT FT/SEC			DOWNRANGE ERROR AT TOUCHDOWN (FT.) -			CROSSRANGE ERROR AT TOUCHDOWN (FT.)		
	$\sigma_H = 24^0$ 25K (SEE NOTE 6)	$\sigma_N = 40^0$	$\sigma_N = 24^0$ 20K	$\sigma_H = 24^0$ 25K	$\sigma_H = 40^0$	$\sigma_N = 24^0$ 20K	$\sigma_H = 24^0$ 25K	$\sigma_N = 40^0$	$\sigma_H = 24^0$ 20K	$\sigma_H = 24^0$ 25K	$\sigma_N = 40^0$	$\sigma_H = 24^0$ 20K	$\sigma_N = 24^0$ 25K	$\sigma_N = 40^0$	$\sigma_B = 24^0$ 20K
NOMINAL	1.1	.6	1.1	NONE	NONE	NONE	6513	6513	6515	-2.5	-2.5	-2.5	0	0	0
TERRAIN MODEL NO. 6	141	2167	136	NONE	6900-7200/13	NONE	6523	6518	6523	86	99	40	0	0	0
-MISALIGNMENT OF -3 MILLIRADIANS	386	2120	391	NONE	9500-4300/42	NONE	6504	6506	6505	-252	-263	-230	0	0	0
X ACCELEROMETER BIAS OF -.75 CM/SEC^2	513	1665	514	NONE	NONE	NONE	6542	6544	6543	-3	14	-2	0	0	0
MIT INITIAL CONDITION NO. 212	-200	-1940	-215	NONE	NONE	NONE	6516	6517	6521	-9150 (5)	-9200 (5)	-9227 (5)	1620 (5)	1622 (5)	1621 (5)
MIT INITIAL CONDITION NO. 1191	229	1607	236	NONE	NONE	NONE	6514	6515	6517	7612 (5)	7600 (5)	7673 (5)	873 (5)	861 (5)	865 (5)
THRUST VARIATION OF +5% TO +4%	25	11	23	NONE	NONE	NONE	6461	6461	6461	-2.5	-2.7	-2.5	0	0	0

(1) ANTENNA ANGLE BELOW HIGH GATE IS 0^0.
(2) σ_N = ANTENNA ANGLE ABOVE HIGH GATE.
(3) RCS FUEL CONSUMPTION DIFFERENCE BETWEEN THE VARIOUS TRAJECTORIES IS LESS THAN ±1 POUND.
(4) POSITIVE ALTITUDE ERROR INDICATES THAT TRUE ALTITUDE IS GREATER THAN COMPUTED ALTITUDE.
(5) LARGE DOWNRANGE AND CROSSRANGE ERRORS RESULT FROM THE PARTICULAR INITIAL CONDITION SETS AND CAN NOT BE CORRECTED BY RADAR DATA. ONLY THE DIFFERENCES BETWEEN $\sigma_N = 24^0$ AND $\sigma_N = 40^0$ RESULTS SHOULD BE CONSIDERED.
(6) FOR $\sigma_N = 24^0$, THE RESULTS ARE SHOWN FOR ALTITUDE INITIALLY UPDATED AT 25,000 FEET AND THEN AT 20,000 FEET. FOR $\sigma_N = 40^0$, THE ALTITUDE IS INITIALLY UPDATED AT 25,000 FEET.

Table 5 ANTENNA ANGLE COMPARISON DATA AND ERROR EFFECTS

requirements. Interaction of the radar with other PGNCS sensor errors, and I.C. errors does not increase the zero Doppler loss of track, considering each error separately at its 3σ value. Thrust variation uncertainty does not deteriorate radar performance within the limits tested. The two terrain models "flown" are within the scope of the radar and PGNCS system operation so that mission and radar specifications are complied with.

Redesignation with the specified circular footprint can be accomplished within the ΔV constraint and without causing additional zero Doppler track loss.

7. General Conclusions. Referring back to Part C, "Statement of the Problem," the answers can now be given. Under the assumed radar configuration and within the limitations of the simulation assumptions and accuracy it can be stated that the radar system operates in compliance with the radar system requirements and, in conjunction with the other systems of the PGNCS, it operates in compliance with mission requirements.

The coupling of radar measurement errors through the control system is negligible. The LR measurement errors allow compliance with the explicit guidance law aim points to a high degree of accuracy, sufficient to meet mission requirements.

The zero Doppler problem is not excessive. Loss occurs at antenna switching, at the start of guidance subphase two and at the hover point.

The effects of the lunar terrain models studied are within the scope of acceptable radar operation. The effects of an increased lunar reflectivity would enhance radar operation by extending upwards the altitude to which acquisition occurs and by increasing the S/N ratio throughout the trajectory.

The effects of thrust variation, gyro and accelerometer errors, and I.C. errors are within the scope of acceptable radar performance. Redesignation of landing sites, as specified in the study, is possible without reducing radar system performance.

There is compliance with

- the touchdown velocity envelope requirement,
- ΔV budget restrictions,
- visibility of the landing site requirements,
- 1/2 mile CEP requirements, and altimeter and velocity data is available at the required altitudes.

III. Guidelines

A. Problem Formulation

The guidelines which were profitably applied to the problem formulation are very close to what is taught in all courses in mathematics: one must know what is given, what is required, and decide what method to apply, then formulate a time-order task list and follow it. Flexibility is neces-

sary to enable one to adjust to problem changes which occur concurrently with its solution.

B. Constructing Math Models

In the case of the landing radar which is a very complex system, it was necessary to evaluate the various error contributing effects and consider only those of major magnitude. It was required to decide what level of error source should be included. This study was made prior to constructing the math model and proved very useful. Basically the ground rule followed was: make the math model in the level of detail and completeness which corresponds with the nature of the problem and the tool (computer complex) to be used for mechanization. This in turn implies an appraisal and sorting of the levels of the actual system.

C. The Study Plan

Again, the rule which was followed was: design your study plan to answer the problem in sufficient but not excessive completeness. To do this required thought, planning and care. Far too often the temptation to make a large number of runs is yielded to, rather than keeping the number in correspondence with the real needs of the problem.

D. Adapting to Design Changes

In the study reported in this paper it has been pointed out in many places that design changes occurred during the study. This is inevitable in the case of a problem of this nature where prototypes are being fabricated and tested along with the analysis. It is possible to be flexible in the constructing of the math model where one knows changes may occur by putting into the equations symbols in place of numerical constants. In the programming, allowance may be made for changes in the integration scheme, and internal logic. This may be aided by carefully annotating the program with clear comments so that a one to one correspondence with the equations may be easily made. Finally in the study plan itself, a certain adaptability can be built in and use made of earlier results to adapt them as much as possible to the new design and rerun only a sufficient number of cases to establish the validity of the adaptation. For example, once it had been demonstrated that the effect of coupling of radar measurement errors with the guidance and control system was minimal with antenna angles of degrees (50.7, -20.4), it was only necessary to run a few cases to establish that this applied equally well to (40, 0) degrees and (24, 0) degrees.

IV. Bibliography

1. Static Error Analysis Report on LM Landing Radar, Ryan Electronics, Ryan Aeronautical Co., Report 53969-50, January 1967.

2. Kriegsman, B. and Sears, N., LEM PGNCS and Landing Radar Operation During the Powered Lunar Landing Maneuvers, MIT Instrumentation Laboratory, E-1982, August 1966.

3. Harrington, R. L. and Zak, D. D., Landing Radar Performance Analysis Program, Ryan Electronics, Ryan Aeronautical Co., Report 53969-34.

4. Sears, N. E., MIT/IL, Hoffman, P. F., GAEC, LEM-PGNCS Guidance Equations for a Nominal Lunar Landing Mission, MIT Instrumentation Laboratory, E-1981, May 1966.

5. Eichler, J., The Dynamic Performance of the LM Landing Radar, LMO-500-460, August 1966.

6. Eichler, J., Landing Radar Math Model, LMO-500-463, July 1966.

7. Lunar Module Primary Guidance, Navigation and Control Subsystem Equipment Performance and Interface Specification, LSP 370-3, 10 March 1966.

8. Eichler, J., Birnbaum, S., The Guidance and Control Verification Study, LED 500-10, March 1966.

9. Design Reference Mission II A, Apollo Mission Planning Task Force MSC, Report PM 3/M-171/66, October 1966.

10. LEM Landing Radar Technical Review, 25 October 1966, MSC Houston Texas, LMN-500-70 published 2 November 1966.

11. Eichler, J. and Gardos, R., Landing Radar Antenna Angle Study, LMD 500-532, 12 January 1967.

12. Eichler, J., Verification of Performance of the LM Guidance and Control Systems - AIAA Flight Test, Simulation and Support Conference, AIAA Paper 67-244.

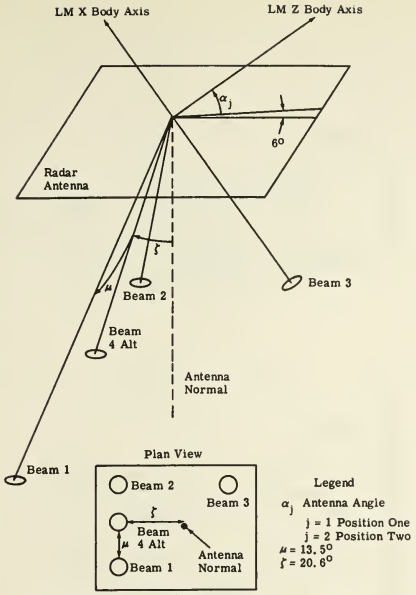


FIGURE 4 - RADAR BEAM GEOMETRY

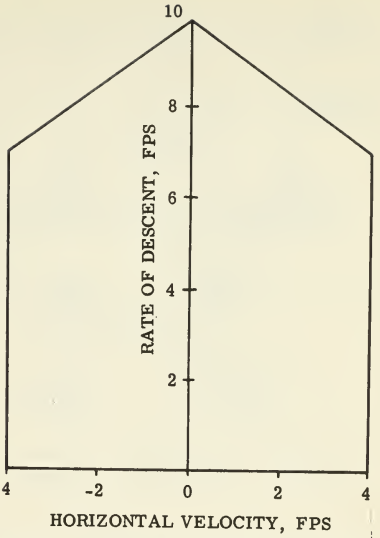


FIGURE 6 - TOUCHDOWN VELOCITY ENVELOPE

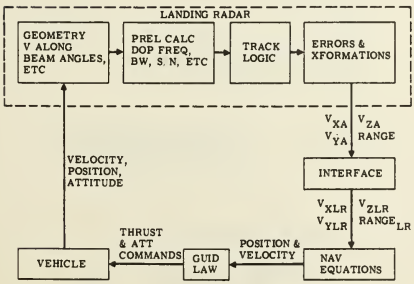


FIGURE 5 - LOGICAL FLOW CHART

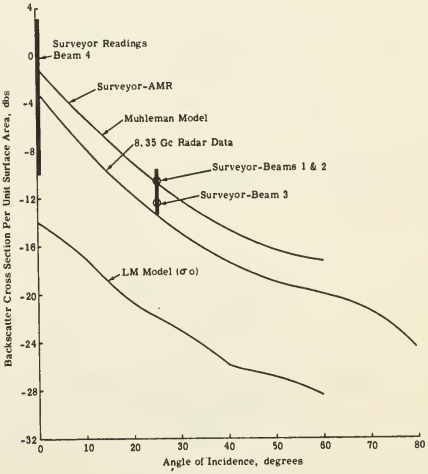


FIGURE 7 - REFLECTIVITY VS ANGLE OF INCIDENCE

67-622

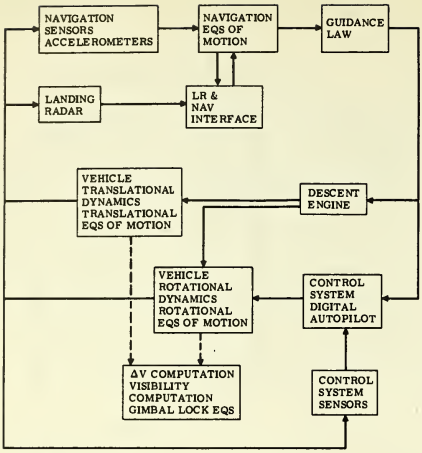


FIGURE 8 - . SIMULATION FLOWCHART

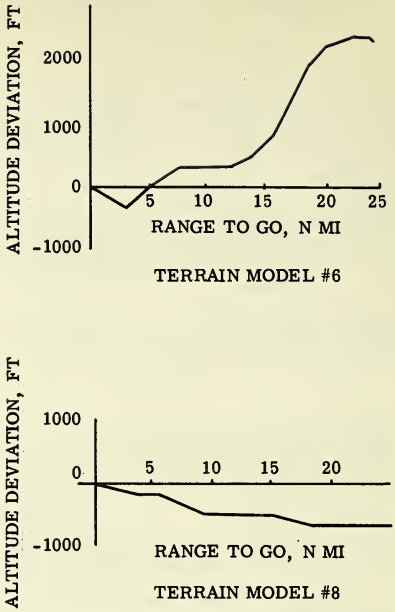


FIGURE 10 - TERRAIN MODELS

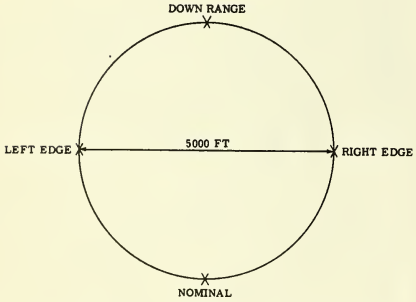


FIGURE 9 - FOOTPRINT OF POSSIBLE LANDING SITES

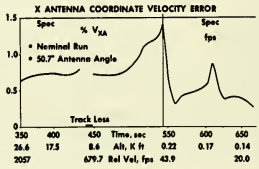


FIGURE 11 - X ANTENNA COORDINATE VELOCITY ERROR

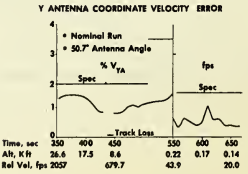


FIGURE 12 - Y ANTENNA COORDINATE VELOCITY ERROR

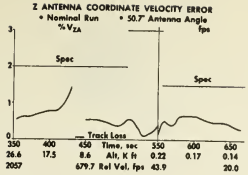


FIGURE 13 - Z ANTENNA COORDINATE VELOCITY ERROR

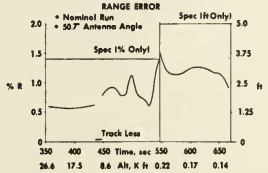


FIGURE 14 - RANGE ERROR

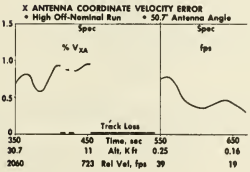


FIGURE 15 - X ANTENNA COORDINATE VELOCITY ERROR

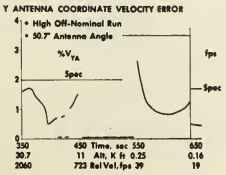


FIGURE 16 - Y ANTENNA COORDINATE VELOCITY ERROR

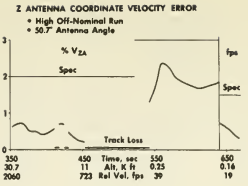


FIGURE 17 - Z ANTENNA COORDINATE VELOCITY ERROR

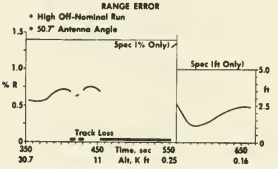


FIGURE 18 - RANGE ERROR

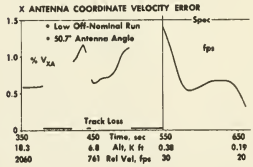


FIGURE 19 - X ANTENNA COORDINATE VELOCITY ERROR

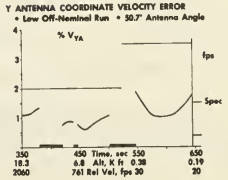


FIGURE 20 - Y ANTENNA COORDINATE VELOCITY ERROR

67-622

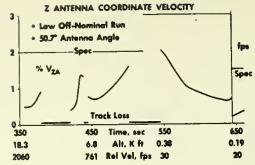


FIGURE 21 - Z ANTENNA COORDINATE VELOCITY ERROR

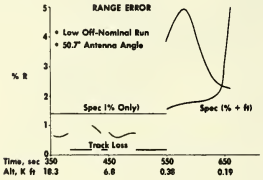


FIGURE 22 - RANGE ERROR

ΔX	4.7 ± 5.6 (FT)
ΔY	-0.7 ± 11.2 (FT)
ΔZ	19.0 ± 2.2 (FT)
$\Delta \dot{X}$	0.68 ± 0.49 (FT/SEC)
$\Delta \dot{Y}$	0.51 ± 2.53 (FT/SEC)
$\Delta \dot{Z}$	3.60 ± 0.65 (FT/SEC)

FIGURE 23 - SENSITIVITY AT HI GATE NAV TO LR (1σ)

	NAV TO LR (1σ)	TRAJ TO LR (1σ)
ΔX	-10 ± 6 (FT)	-10 ± 6 (FT)
ΔY	-96 ± 171 (FT)	-94 ± 167 (FT)
ΔZ	41 ± 21 (FT)	232 ± 6 (FT)
$\Delta \dot{X}$	-0.12 ± 0.19 (FT/SEC)	-0.04 ± 0.2 (FT/SEC)
$\Delta \dot{Y}$	-0.16 ± 0.54 (FT/SEC)	0.50 ± 1.6 (FT/SEC)
$\Delta \dot{Z}$	-0.28 ± 0.23 (FT/SEC)	-1.62 ± 1.6 (FT/SEC)

FIGURE 24 - SENSITIVITY AT LO GATE

	T_{LOSS} (SEC)	T_{REGAIN} (SEC)	SHIFT FROM ZERO ERRORS, SEC
ZERO ERRORS	435.85	441.85	-
DETERMINISTIC ERRORS ONLY	437.85	443.85	2
INSTALLATION ERRORS (1σ) ONLY	437.65	443.65	1.8
FLUCTUATING ERRORS (1σ) ONLY	439.35	445.35	3.5

$\Delta T = 6.00$ FOR ALL CASES

FIGURE 25 - TRACK LOSS

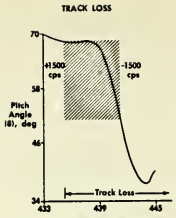


FIGURE 26 - TRACK LOSS

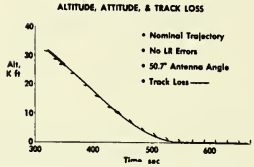


FIGURE 27 - ALTITUDE, ATTITUDE, AND TRACK LOSS

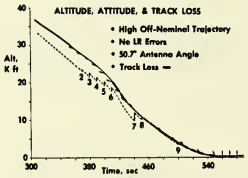


FIGURE 28 - ALTITUDE, ATTITUDE, AND TRACK LOSS

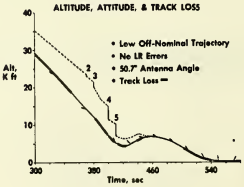


FIGURE 29 - ALTITUDE, ATTITUDE, AND TRACK LOSS

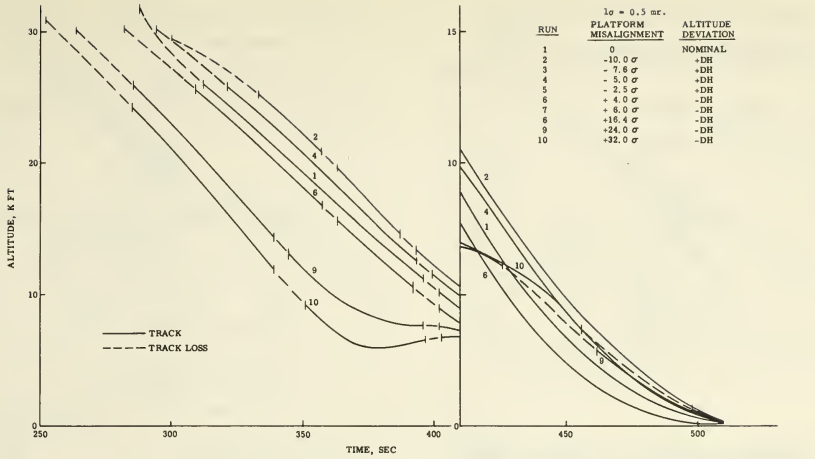


FIGURE 30 - A FAMILY OF ERROR TRAJECTORIES DUE TO PITCH MISALIGNMENT

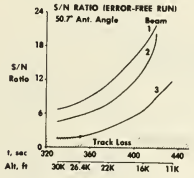


FIGURE 31 - S/N RATIO (ERROR-FREE RUN)

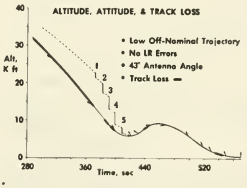


FIGURE 33 - ALTITUDE, ATTITUDE, AND TRACK LOSS

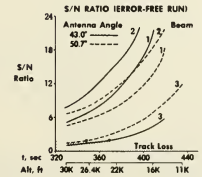


FIGURE 32 - S/N RATIO (ERROR-FREE RUN)

RESULTS	Antenna Angles	
	40.0	50.7, −20.4
A. Nominal reflectivity		
Accuracy compliance	Yes	Not tested
Alt track acquisition	35K Ft	40K
Vel track acquisition	19.8K Ft	11.6K Ft
Subsequent track loss A&V	Only at	34K-11.6K Ft
	Δ change	10K-7K Ft
S/N ratio	Superior	Inferior
B. Assume +5 db reflectivity		
•Alt track acquisition	> 40K Ft	40K
•Vel track acquisition	31.4K Ft	34K
•Subsequent track loss A&V	Only at	29K-14.5 K Ft
	∡ change	10K-6.9K Ft

Note:
10MV noise, 10 sec alt sampling

FIGURE 34 - RESULTS

67-622

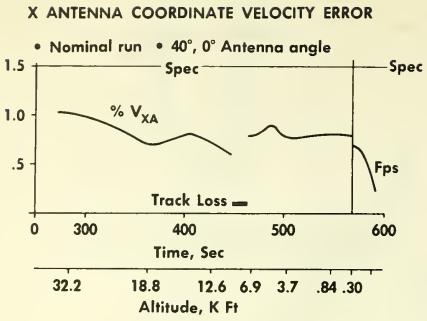


FIGURE 35 - X ANTENNA COORDINATE VELOCITY ERROR

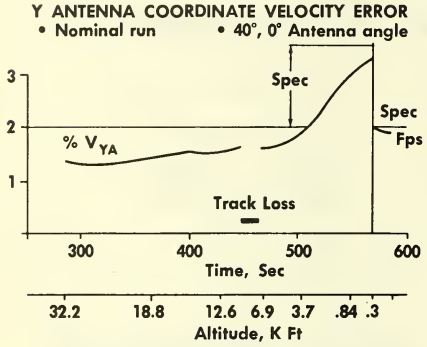


FIGURE 36 - Y ANTENNA COORDINATE VELOCITY ERROR

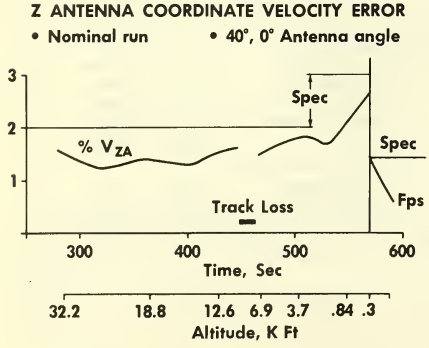


FIGURE 37 - Z ANTENNA COORDINATE VELOCITY ERROR

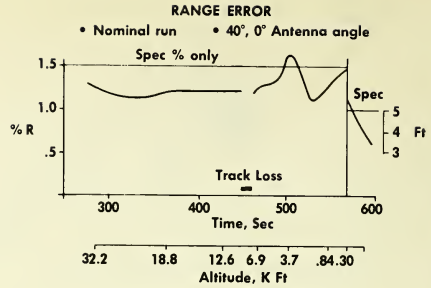


FIGURE 38 - RANGE ERROR

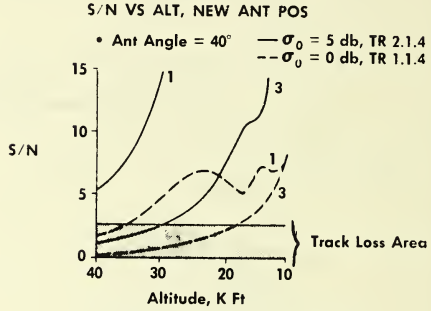


FIGURE 39 - S/N VS ALTITUDE, NEW ANT POS

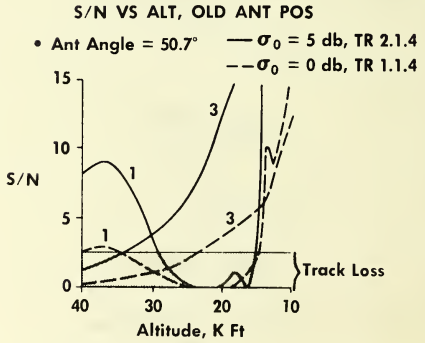


FIGURE 40 - S/N VS ALT, OLD ANT POS

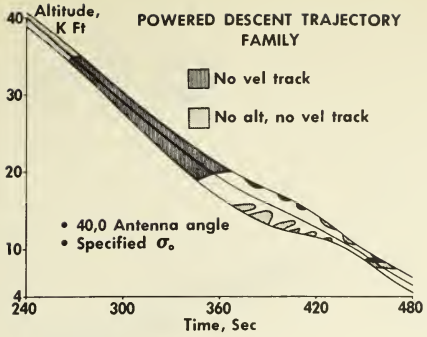


FIGURE 41 - POWERED DESCENT TRAJECTORY FAMILY

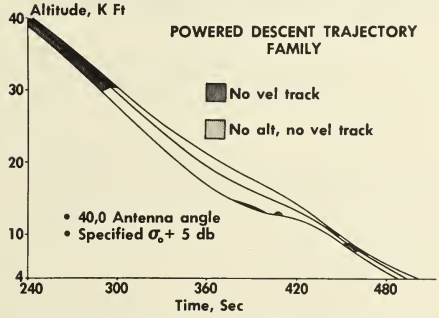


FIGURE 42 - POWERED DESCENT TRAJECTORY FAMILY

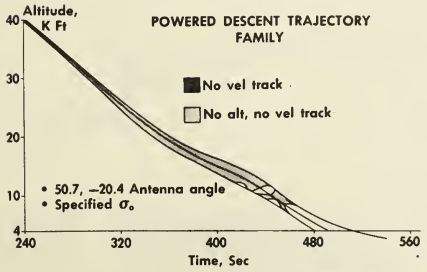


FIGURE 43 - POWERED DESCENT TRAJECTORY FAMILY

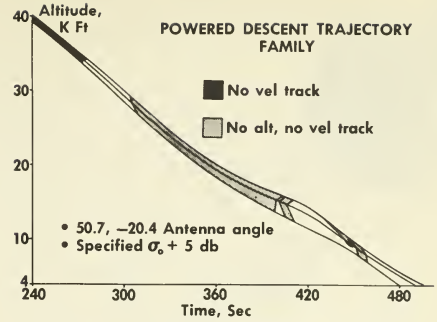


FIGURE 44 - POWERED DESCENT TRAJECTORY FAMILY

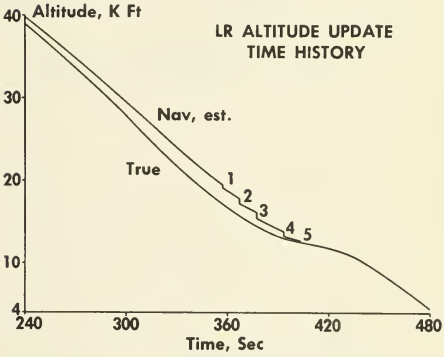


FIGURE 45 - LR ALTITUDE UPDATE TIME HISTORY

Nominal Trajectory (LPD Guidance Law)

Altitude (Ft)	Local Pitch $\measuredangle$ (Degrees)	Velocity (Fps)		Thrust (Lbs)
		$\dot{X}$	$\dot{Z}$	
32189	73.8	-44	2676	10016
24915	68.8	-98	2147	10064
18794	65.7	-92	1598	10112
14960	70.7	-65	1040	10153
12574	66.6	-114	831	5384
10431	65.0	-140	700	5444
* 8618	39.8	-150	605	5561
6908	40.5	-129	532	5387
5023	40.7	-103	437	5133
3696	40.3	-84.3	356	4891
1579	36.4	-49.1	191.4	4279
842	32.0	-33.2	115.2	3918
301	23.6	-16.4	43.7	3511
118	12.4	- 2.07	3.79	3243

*Switch to Positon Two

FIGURE 46 - NOMINAL TRAJECTORY (LPD GUIDANCE LAW)

67-622

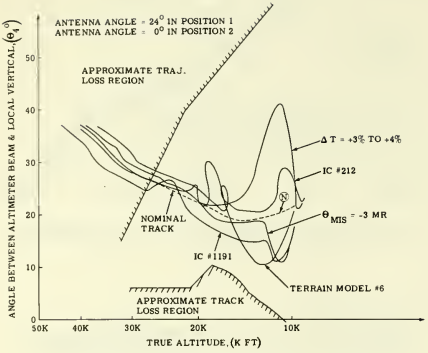


FIGURE 47 - ERROR CASES, TRACK DIAGRAM $\alpha = 24^{\circ}$

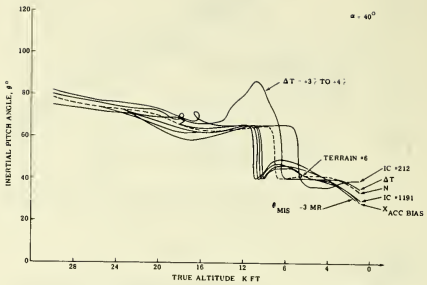


FIGURE 49 - PITCH VS ALTITUDE $\alpha = 40^{\circ}$

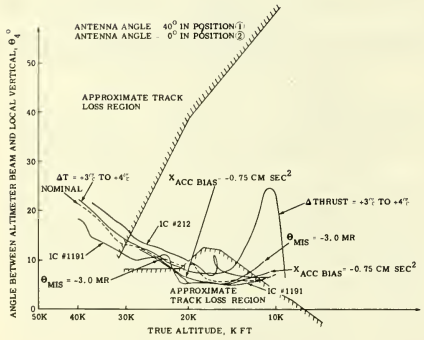


FIGURE 48 ERROR CASES, TRACK DIAGRAM $\alpha = 40^{\circ}$

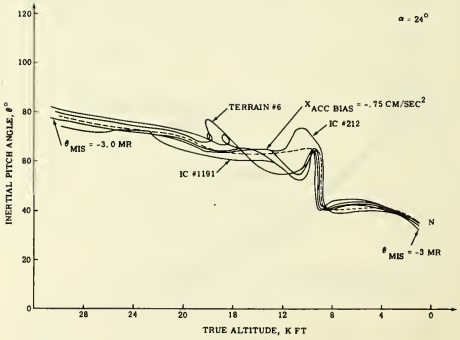


FIGURE 50 - PITCH VS ALTITUDE $\alpha = 24^{\circ}$

No. 67-623

THE REGION OF KALMAN FILTER CONVERGENCE FOR SEVERAL AUTONOMOUS NAVIGATION MODES

by

N. F. TODA, F. H. SCHLEE,
and
P. OBSHARSKY
International Business Machines Corporation
Endicott, New York

AIAA Paper
No. 67-623

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

8.02, 8.07

-- NOTES --

THE REGION OF KALMAN FILTER CONVERGENCE FOR SEVERAL AUTONOMOUS NAVIGATION MODES

N. F. Toda, Senior Engineer
F. H. Schlee, Advisory Engineer
P. Obsharsky, Staff Engineer

International Business Machines Corporation
Federal Systems Division, Space Systems Center
Endicott, New York

Abstract

Autonomous navigation using a Kalman filter requires an initial estimate of six orbit parameters, an initial estimate of the six by six covariance matrix and an a priori estimate of the accuracy of the measurement data. When errors in the initial estimate are too large the Kalman filter fails to converge. Errors are not reduced as more and more data is processed; useful navigation accuracies may never be achieved.

The region of convergence is shown to be larger for autonomous navigation using a star tracker and a horizon sensor than for navigation using a star tracker and a landmark tracking telescope. In using stellar referenced landmark tracking data, the region of convergence is larger when unknown landmarks are used than when known landmarks are used. Within similar types of autonomous navigation systems, there is an inverse relationship between accuracy and the region of convergence; the more accurate the navigation system the smaller the region of convergence.

The region of convergence can be increased by choosing the initial covariance matrix smaller than the expected actual errors or by employing an a priori instrument error covariance matrix which is larger than the expected instrument errors.

Convergence has been obtained starting from an initial error of 1,000 miles down range and 100 miles cross range using a star tracker-horizon sensor system.

These tolerances imply that the Kalman filter could be initialized by an astronaut making an unaided visual observation of the ground.

Introduction

Autonomous navigation is the process of determining the orbit of a spacecraft from on-board measurements. With both data measurement and computation functions performed on board the spacecraft, the navigation process is completely independent of ground support facilities. The inherent flexibility and quick reaction to changes in mission plan make autonomous navigation highly desirable for long-duration space missions, particularly the low earth orbital, manned missions. This paper considers autonomous navigation in a low earth orbit environment. A Kalman filter is used to process a sequence of stellar referenced or earth referenced measurements. The latest measurement data and the current best estimate of the orbit are processed by a Kalman filter to obtain a differential correction to the estimated orbit, i.e., a more accurate estimate of six orbit parameters. The orbit parameters, also referred to as state variables, are the three components of spacecraft position and the three components of spacecraft velocity.

The Kalman filter is a recursive minimum variance estimator for linear problems. Utilization of the technique in space navigation therefore requires that the orbit determination problem be linearized about the current best estimate of the trajectory. Since the six state variables describing the orbit satisfy nonlinear differential equations and since the measurements are usually nonlinear functions of the state variables, the initial estimate of the state vector must be close to the true state to permit the linear recursive process to converge. If errors in the initial estimate are too large, the Kalman filter may not function properly and the process may not converge. Instead, errors may become very large and useful navigational accuracies may not be achieved even though many navigational fixes have been processed.

Several different forms of filter stability have been discussed in recent publications. Bellantoni[1] was concerned with the loss of numerical significance in the computation of the state covariance matrix which is employed to compute the gain of the filter. Nonlinearities were excluded from Bellantoni's work; only the round off effects in a finite word length computer were included. In discussing measurement function nonlinearity, Denham and Pines[2] have stated, "The essential difficulty is not that the estimated state, after processing a data point, is seriously affected by nonlinearity, but rather that the calculated covariance matrix of estimation errors becomes inaccurate, thus making future data processing nonoptimal."

Long-term stability problems have also been discussed. Schlee[3] described the application of Pine's[4] modified Kalman filter to controlling long-term divergence, i.e., a divergence of the estimated state from the true state which occurs after the filter has been operating properly for some time. This phenomenon is caused by round off in the finite word length computer and by errors in the assumed mathematical model for which the filter is designed.

67-623

This paper considers the initial convergence of the Kalman filter: how large can the initial errors be before the Kalman filter fails to function? What parameters effect the region of convergence and how may the region be enlarged? Results presented here describe both the region of convergence and the relative degree of long term stability for five modes of autonomous navigation. These five modes consist of different combinations of the navigation sensors: star trackers, horizon sensor, altimeter, and a telescope for tracking terrestrial landmarks are combined to form several different autonomous navigation modes. Both known landmark tracking and unknown landmark tracking are included.

The Operation of the Kalman Filter

The principal operations performed in the Kalman filter are diagrammed in Figure 1. The filter generates $\hat{\underline{X}}$ (k) a new estimate of the state of the spacecraft by adding $\delta \underline{x}$, a correction or update, to the previous estimate. Spacecraft state is a six vector, each component of which is an orbit parameter. The update is formed by multiplying the observation deviation $\delta A = \tilde{A} - \hat{A}$ by a weighting matrix $\underline{W}$ (k). The observation δA in turn is the difference between the actual measurement $\tilde{A}$ and the estimated or predicted value, $\hat{A}$. The relationship between the estimate of the state $\hat{\underline{X}}$ (k) after processing the k^{th} measurement and the best estimate before processing the measurement is given by

$$\hat{\underline{X}}(k) = \hat{\underline{X}}(k-1) + \underline{W}(k)\left[\tilde{A}(k) - \hat{A}(k)\right] \qquad (1)$$

The weighting or gain matrix $\underline{W}$ (k) is given by

$$\underline{W}(k) = \underline{P}(k-1)\,\underline{H}^{T}(k)\left[\underline{H}(k)\,\underline{P}(k-1)\,\underline{H}^{T}(k) + Q\right]^{-1} \qquad (2)$$

where $\underline{H} = \dfrac{\partial A}{\partial \underline{x}}$, a six-vector, is the gradient of the observation with respect to the state variables. The row vector $\underline{H}$ is sometimes called the gradient vector or the observation matrix. The scalar Q is the variance of the instrument noise errors. The expression relating the updated six by six covariance matrix $\underline{P}$ (k) to the covariance matrix before the update $\underline{P}$ (k-1) is:

$$\underline{P}(k) = \underline{P}(k-1) - \underline{W}(k)\,\underline{H}(k)\,\underline{P}(k-1) \qquad (3)$$

Each time an observation is processed, the covariance matrix is updated. Note that the covariance matrix satisfies a recursion relation, equation (3), as does the estimated orbit, equation (1).

Both recursion relations require initial input values. Thus, to begin navigating with a Kalman filter one must input an estimate of the elements of the state vector $\hat{\underline{X}}$ (o) and also the elements of the initial covariance matrix, $\underline{P}$ (o).

In practice, a modification to the Kalman filter has been found to be necessary to prevent filter divergence caused by modeling and computer round-off errors[3]. As shown in Figure 1, filter modification is an additive term to the convariance matrix.

Figure 2 shows the convergent behavior of the Kalman filter for a known landmark navigation mode. The initial position error of 26,000 feet is reduced to an RSS error of 221 feet after processing data from ten known landmarks over a three-orbit interval. On the other hand, Figure 3 shows a divergent behavior caused by quadrupling the initial errors and the initial covariance matrix. Both figures also show the relationship of RSS and RMS errors, where:

- Root Sum Squared (RSS) error is the length of the vector difference between the true spacecraft position and the estimated spacecraft position.

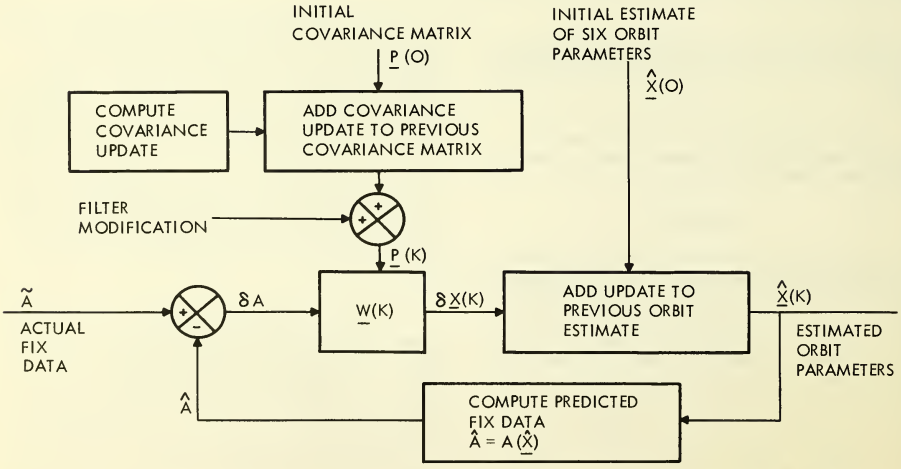


Figure 1. Kalman Filter

- Root Mean Squared (RMS) error is the square root of the trace of the position covariance matrix.

$$RMS = \sqrt{\sigma_R^2 + \sigma_T^2 + \sigma_N^2}$$

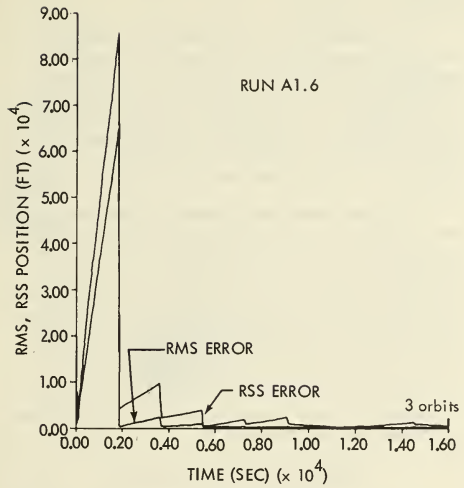


Figure 2. Convengent Behavior of the Kalman Filter for a Known Landmark

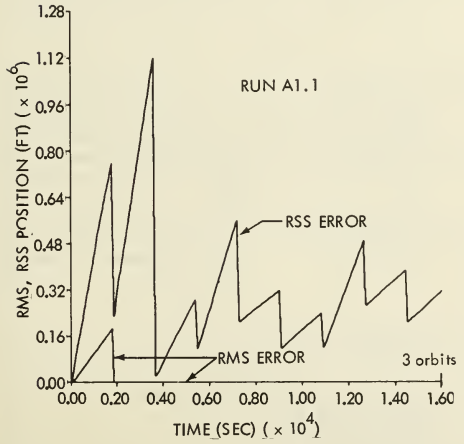


Figure 3. Divergent Behavior of the Kalman Filter for a Known Landmark

Autonomous Navigation

Five modes of autonomous navigation are considered:

Mode A – Known Landmark

The line of sight to a known terrestrial landmark is measured in a stellar orientation frame. Two automatic star trackers and a gimballed telescope for tracking landmarks, i.e., earth fixed points, are the navigation sensors employed. The state vector has nine elements which includes the three position coordinates of the landmark in addition to the six orbit parameters.

Mode B – Unknown Landmark

The line of sight to an unknown terrestrial landmark[5] is measured in a stellar orientation frame. Sensors used are the same as in Mode A. The difference between Modes A and B lies in the landmarks: for Mode A the position coordinates are accurately known prior to the fix taking operations, for Mode B the landmark coordinates are initially unknown. The first sighting to an unknown landmark is utilized to obtain a preliminary estimate of the location of the landmark. Subsequent sightings refine the estimate of the orbit parameters and the landmark coordinates. The same nine element state vector is estimated as in Mode A.

Mode C – Star Vertical

The direction of the local vertical is measured in a stellar orientation frame. The sensors used are two star trackers and a horizon sensor.

Mode D – Star Vertical with Altimeter

Mode D measurements are the same as Mode C with the addition of an altitude measurement. The sensors used are the same as in Mode C with the addition of a radar altimeter.

Mode E – Landmark-Vertical

The orientation of the line of sight to a known landmark is measured in a local vertical orientation frame. Equipment used is the horizon sensor and the landmark sighting telescope. A ten-element state vector is estimated. In addition to the six orbit parameters and three landmark coordinates, Mode E estimates horizon sensor tilt.

The relative navigational accuracies achievable in these modes is dependent upon the orientation source (stellar or horizon) and the object tracked (landmark or local vertical). A stellar orientation frame is superior to a local vertical frame because of the order of magnitude better pointing precision of a star tracker relative to a horizon sensor. Thus, Modes A and B are more

accurate than Mode E. A known landmark is more useful for precise autonomous navigation than an unknown landmark. Thus Mode A is more precise than Mode B. Tracking a landmark is more useful than tracking the local vertical. Thus Modes A and B are more precise than Mode C. Addition of more data is always beneficial: adding altimeter data makes Mode D more precise than Mode C. These statements are, of course, all dependent upon the observation schedule used and it would be possible for instance, to choose observation schedules so that Mode C would be more accurate than Mode A. However, for reasonable observation schedules, the most accurate autonomous navigation mode is A, followed by B, E, D and C in order of decreasing accuracy.

Simulation Study

Navigation system behavior was studied by simulating the entire navigation process on a digital computer. An environment simulation generated the actual spacecraft position, landmark location and also star sighting data. The Kalman filter simulation employs a single precision word length (IBM 7094) to process the simulated fix data into an estimate of the spacecraft parameters. Random number generators simulated the noise-like errors that occur in making a navigational fix.

Navigation studies were performed for a spacecraft in a 160 nautical mile circular earth orbit. The set of "standard" parameters describing observation schedules, instrument accuracies, and landmark accuracies is shown in Table 1. Random errors noted in this table are simulated as zero-mean Gaussian random sequences. Random measurement errors are assumed to be statistically independent. The bias vertical misalignment in E is a zero mean random variable which is constant for a single landmark tracking period but varies randomly from landmark to landmark. This error behavior reflects the fact that horizon sensor errors are caused by anomalies in the horizon which cause a constant "tilt" error during the short two-minute period that any single landmark is tracked.

The relative geometry involved in a landmark tracking sequence is shown in Figure 4. Landmarks are acquired when the line of sight to the landmark is about 45 degrees from the vertical. During a landmark tracking operation, the telescope tracks the landmark either manually or automatically. Gimbal angles which define the angular orientation of the telescope to the reference frame are measured at intervals of 10 seconds.

Analysis studies used uniformly spaced landmarks: i.e., for Mode A, one known landmark is tracked each 120 degrees of orbit travel, for Mode B, one unknown landmark is tracked each 40 degrees of orbit travel. Landmark out-of-plane orientation was varied throughout. The larger number of unknown landmarks reflects the ease of acquiring unknown landmarks which can be any conveniently located, easily distinguishable, terrestrial feature. Weather and lighting conditions can make recognition of a known landmark difficult. For these reasons fewer known landmarks than unknown landmarks are available in any given period.

Characterization of Initial Errors

Initial estimates of six orbit parameters and a six by six covariance matrix must be supplied to initialize the Kalman filter. To facilitate the parametric study of filter convergence, the choice of initial errors was

Table 1

AUTONOMOUS NAVIGATION SYSTEM PARAMETERS

Navigation System	Observation Schedule	One Sigma Pointing Accuracy	One Sigma Landmark Accuracy
Mode A Star Tracker Landmark Telescope	9 sets of pitch and roll angle sightings over each landmark 3 landmarks each orbit	40 $\widehat{\text{sec}}$. random error in telescope pointing	Known landmark uncertainty is 300 feet in each axis
Mode B Star Tracker Landmark Telescope	9 sets of pitch and roll angle sightings over each landmark 9 landmarks each orbit	40 $\widehat{\text{sec}}$. random error in telescope pointing	Uncertainty in unknown landmark height above sea level is 300 feet
Mode C Star Tracker Horizon Sensor	1 pair of star-vertical angle measurements each 10 min.	0.1 degree random error in vertical (horizon sensor) pointing	Not applicable
Mode D Star Tracker Horizon Sensor Radar Altimeter	1 pair of star-vertical angle measurements each 10 min. 1 altimeter measurement each 10 minutes	0.1 degree random error in vertical (horizon sensor) pointing. 300 foot altimeter (geoid) accuracy	Not applicable
Mode E Horizon Sensor Landmark Telescope	9 angle measurements: angle between telescope line of sight and the vertical (horizon sensor) reference 3 landmarks each orbit	0.1 degree random error in vertical (horizon sensor) pointing. 0.1 degree bias error in vertical (horizon sensor) pointing	Same as A

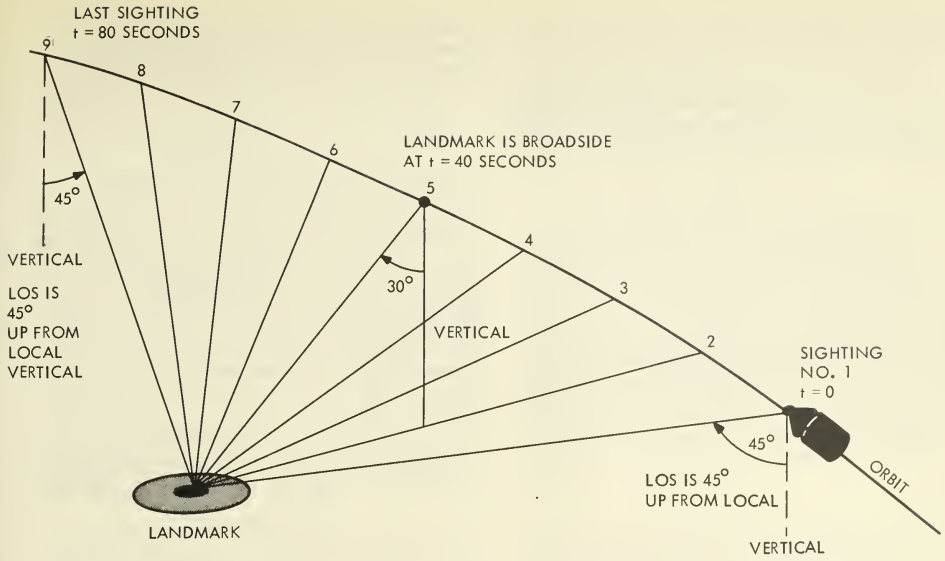


Figure 4. Landmark Tracking Geometry

restricted in the following manner: the magnitude of initial position errors in three orthogonal coordinates is S, the magnitude of the velocity errors is ωS where ω is the orbital rate in radians per second. Initial covariance matrices are always assumed to be diagonal with diagonal terms of S^2 in the three position coordinates and $\omega^2 S^2$ in the three velocity coordinates.

With these restrictions, the choice of the S parameter fixes all elements of the covariance matrix and also all magnitudes of the initial estimation errors. Signs in the initial errors are left unspecified. In order to determine the region of convergence for a given navigation scheme, an initial choice of the S parameter is made. Then 32 simulations are performed, each one starting with a different combination of signs of the initial errors. Symmetry across the orbital plane is used to reduce the 64 possible combinations of signs of the initial errors down to 32.

The simulation study proceeded by making an initial set of 32 runs and then choosing a worst-case combination of initial error signs. This worst-case combination was then used in simulations where the S parameter was increased until the limit of convergence was determined. This limit is characterized as the S parameter for which the navigation process evidenced a transition from convergent behavior to non-convergent behavior.

Evaluation Criteria

The determination of convergent or non-convergent navigation behavior was based upon plots of the RMS and RSS error history. Figures 2 and 3 show respectively a convergent and a non-convergent behavior. For Figure 2 the initial errors are said to be within the convergent region; for Figure 3 the initial errors are said to be outside the convergent region.

The RSS error, based on the difference between true and the estimated spacecraft position, is a measure of the accuracy actually achieved by the Kalman filter estimation process. The RMS error, based upon the estimation error covariance matrix, is an ensemble statistic derived from a linearized model of the navigation process and is a measure of the theoretical or predicted accuracy. When round off effects and nonlinear effects are small, the RMS error is a good estimate of the RSS error. However, when large nonlinear or round off effects are present, the RSS error will be significantly larger than the RMS error.

A given simulation was judged to be convergent if after an initial transient period the RSS position error either showed a general tendency to decrease or to oscillate in the neighborhood of the RMS error. As an aid to the process of deciding whether a given simulation was convergent and to reduce the number of graphs required to determine the region of convergence, three numbers were employed: the RSS position error after three orbits, the ratio of the RSS position error to the RMS error and a sum criterion. The sum criterion is an average of the RSS errors occurring in a single run. It is computed by adding up all the RSS errors present after each landmark tracking operation (or after each pair of star vertical measurements) and dividing by the number of

landmarks (or number of pairs of star vertical measurements).

A graph of the RSS sum criterion versus S was employed to indicate significant departures from linearity. The S value defining the region of convergence is slightly larger than the point at which the nonlinearity becomes evident.

Results

A sequence of simulation runs was performed for each navigation mode to determine which combination of signs of the initial position and velocity errors resulted in the slowest filter convergence. After this combination had been determined, a sequence of simulations was performed with increasing magnitudes of the initial errors. When the limit of the convergence region had been determined, various other filter or observation schedule parameters were varied. The following discussion treats the navigation modes one at a time, describes the region of convergence and effect of certain parameters upon the convergence.

Mode A

Simulation data from a sequence of runs using the worst case sign combinations is shown in Table 2. Runs made with progressively increasing values of the initial error are seen to result in larger actual errors (RSS errors) and in a larger sum criteria. Because of the large number and high accuracy of the measurements the RMS error is not affected by the variation in the initial covariance matrix from a 30,000 foot level to a 55,000 foot level. Thus, if there were no nonlinear effects, increasing initial error levels would have no effect upon navigation accuracy. The contrary, however, is true since the table indicates a large change in RSS value as S is increased. It can therefore be concluded that increasing initial error levels from 30,000 feet to 55,000 feet introduces nonlinear effects which significantly degrade the actually achieved navigation accuracy.

Mode A Convergence Region

On the basis of data shown in Table 2, from inspection of the entire time history of RSS and RMS errors, Figure 5, and from the graph of the RSS sum criterion versus S, shown in Figure 6, the limit of the region of convergence was identified as S = 35,000 feet.

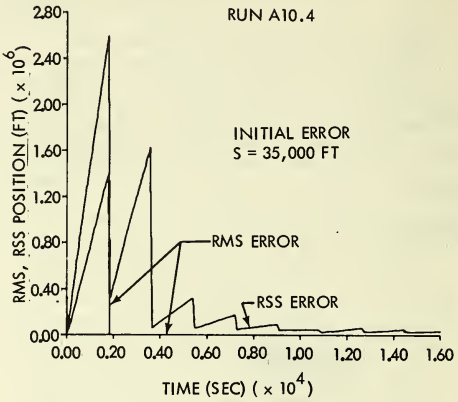


Figure 5. Mode A Limiting Case

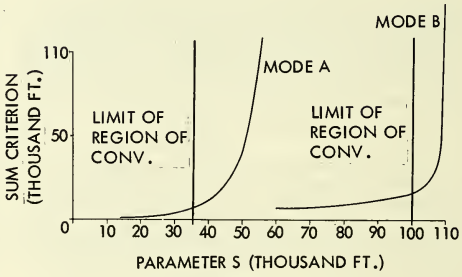


Figure 6. Nonlinear Effects in Modes A and B

Table 2

SEQUENCE OF RUNS USING WORST-CASE SIGN COMBINATIONS

Run	S Initial Error (Feet)	Final RMS Error (Feet)	Final RSS Error (Feet)	RSS/RMS	Sum Criterion (Feet)
A 10.5	30,000	263	1,329	5.0	4,634
A 10.4	35,000	263	2,178	8.3	6,958
A 10.3	40,000	263	3,727	14.0	11,394
A 10.2	45,000	263	7,347	28.0	21,280
A 10.0	50,000	263	21,700	82.0	44,900
A 10.1	55,000	263	88,968	337.0	99,454

Figure 5 shows the time history of RSS and RMS errors for run A 10.4. Examination of the plot shows that subsequent to an initial transient period, the average slope of the RSS curve is negative. To complement the conclusion, an inspection of Figure 6 indicates that a marked deviation in linearity occurs at S = 35,000 feet.

The capability to recognize, acquire and track known landmarks only under daylight conditions, is a variation of Mode A navigation. A series of runs was performed to determine the region of convergence for day only tracking. The results of this investigation, summarized in Table 3, show that the region of convergence is reduced from 35,000 to 30,000 feet by restricting landmark tracking to day only operations. The slight effect on the region of convergence of day only tracking is due to the fact that the navigation process begins at entry into a daylight region (dawn) and the convergence region is largely determined by the initial observation schedule. Since the day-night and day-only schedules are identical during the first half orbit, and since a major reduction of navigation errors occurs during this first half orbit, the region of convergence of the day-only schedule should be near that of the day-night schedule.

Increasing the a priori variance of data measurement errors, the term Q in equation (2), improves the rate of convergence. A series of runs was made in which the actual measurement accuracy (Sigma) was held constant but the Q term in the filter was varied. The results summarized in Table 4 indicate that the region of convergence can be increased to 40,000 feet by making the Q term more than an order of magnitude larger than Sigma, the actual data measurement error level. Day-night landmark schedules were used for this analysis. The artificial increase of the Q parameter artificially increases the RMS error statistics computed by the Kalman filter. However, the actual RSS errors and the Sum Criterion are reduced. Denham[5] noted the beneficial effects of increasing the Q parameter.

Other parameters included in the Mode A convergence study were the effects of errors in the gravity and drag models used for the autonomous navigation, the dynamic biases[3]. Also included was a modification to the Kalman filter* to account for the effects of digital computer round off. These studies showed little change in Mode A convergence: presence of dynamic biases did not significantly impede convergence; addition of filter modifications did not significantly enlarge the region of convergence. However, long term filter stability is significantly degraded by the presence of the dynamic biases[3].

* The machine noise modification[3] adds the diagonal matrix formed from the state variables

$$\underline{P}(k) = 10^{-2\alpha} \begin{bmatrix} x_1(k)^2 & & O \\ & \ddots & \\ O & & x_6(k)^2 \end{bmatrix}$$

to the covariance matrix after each update. α is a parameter which was varied during the study. Investigations were made for α's between 5 and 8.

Table 3

EFFECT OF DAY ONLY LANDMARK TRACKING – MODE A

Run	LM Schedule	S Initial Error (Feet)	Final RMS Error (Feet)	Final RSS Error (Feet)	RSS / RMS	Sum Criterion (Feet)
A 10.4	Day – Night 3 LM/Orbit	35,000	263	2,178	8.3	6,958
ADO.1	Day Only 2 LM/Orbit	35,000	315	9,268	29.4	10,846
ADO.2	Day Only 2 LM/Orbit	30,000	315	3,819	12.1	6,643

Table 4

EFFECT OF VARYING Q PARAMETER FOR MODE A

Run	S Initial Error (Feet)	Q Equation 2	Final RMS Error (Feet)	Final RSS Error (Feet)	RSS / RMS	Sum Criterion (Feet)
A10.3	40,000	$Q=(Sigma)^2$	263	3,727	14.15	11,395
AQ10.3	40,000	$Q=(4Sigma)^2$	403	1,418	3.5	6,073
AQ10.5	40,000	$Q=(16Sigma)^2$	1,260	1,200	0.95	5,312

Mode B

Mode B has nine stellar referenced, unknown landmark tracking operations each orbit instead of the three known landmarks tracked each orbit in Mode A. Figure 6 shows the nonlinear effects to be noticeable at S = 90,000 feet. The large increase in the values shown in Table 5 for S = 110,000 feet indicates that the boundary of the region of convergence occurs at S = 100,000 feet. The navigation process was strongly divergent for S = 120,000 feet.

Reducing the number of landmarks tracked by eliminating measurements during the night time portions of the orbit has a very slight effect on the region of convergence. As in the case of Mode A, the major reduction in the RSS position occurs in the first half orbit where both day-night and day only landmark schedules are the same. Thus, the elimination of the night time measurements has little effect on the region of convergence. Tracking only three landmarks per orbit as in Mode A drastically decreases the region of convergence to 40,000 feet. Comparison of runs A10.3 and BDR1.4 in Tables 2 and 5 shows the unknown landmark navigation to have a smaller RSS error than the known landmark mode. The landmark schedules for both runs are identical. The sum criterion, however, is smaller for Mode A, the known landmark mode. The larger sum criteria in Mode B is caused by the poorer initial transient. Perusal of the time histories of the RSS position error shows that after the third landmark Mode B is better than Mode A. The region of Mode B convergence can be increased by artifically increasing the instrument noise parameter Q. Increasing the Q parameter by a factor of four results in an increase of the Mode B convergence region to 130,000 feet.

The effect of degraded sensor accuracy was investigated by increasing both the Q parameter and the standard deviation of random measurement errors from the nominal level of 40 arc seconds to 120 arc seconds. The region of convergence was found to be 140,000 feet.

Unknown landmarks are points for which there is no a priori knowledge of latitude or longitude. There is, however, some a priori knowledge of the landmark altitude above sea level. The nominal run series shown in Table 5 assumes an a priori landmark altitude knowledge of ±300 feet (one sigma). Increasing this uncertainty to ±3000 feet (one sigma) has a negligible effect upon the region of convergence. An uncertainty of ±10,000 feet in landmark altitude causes a 20 percent decrease in the region of convergence.

The effect of dynamical biases, i.e., an uncertainty in the gravitational constant of 1 part in 100,000 coupled with a 25 percent uncertainty in the drag acceleration has negligible effect on the region of convergence. Dynamical biases do, however, have a noticeable effect on the long term stability of the estimated state vector, though not nearly as pronounced as for Mode A.

Artificially increasing the initial covariance matrix for S = 100,000 feet by a factor of four causes divergence. Table 6 shows the beneficial effect of decreasing the initial covariance matrix by a factor of four in runs BCV 1, 2 and 4.

Table 6

THE EFFECT OF AN ARTIFICIAL DECREASE IN THE INITIAL STATE COVARIANCE MATRIX

Run	S		Final RSS Error (Feet)	Final RSS/RMS	Criterion
	Initial Error (Feet)	Initial Covar. Matrix (Feet)			
B1.01D	100,000	100,000	2,017	1.752	16,203
BCV 2	100,000	25,000	496	0.437	10,413
BCV 1	110,000	27,500	976	0.847	13,432
BCV 4	120,000	30,000	2,402	2.087	19,078

Mode C

Autonomous navigation using a horizon sensor and a star tracker has a very large region of convergence. At a value of S equal to 600,000 feet, the convergence of the system is rapid and does not appear to be near the limit of convergence. Further search for the boundary of the convergence region required a change from the technique in which initial errors and the initial covariance matrix are keyed to the S parameter. At S equal to 600,000 feet the one sigma radial position error of 600,000 feet is already somewhat unrealistic for a 160 nautical mile orbit. Instead, a sequence of simulations was performed in which only the down range position was varied.

Other initial errors were fixed at values shown in Table 7.

Table 5

MODE B CONVERGENCE

Run	LM Schedule	S Initial Error (Feet)	Final RMS Error (Feet)	Final RSS Error (Feet)	RSS/RMS	Sum Criterion (Feet)
B8 1	Day Night	50,000	1,155	605	0.524	5,683
B1.01	Day Night	60,000	1,155	548	0.474	6,428
B1.01A	Day Night	70,000	1,155	470	0.407	7,283
B1.01B	Day Night	80,000	1,154	429	0.371	8,807
B1.01C	Day Night	90,000	1,153	757	0.657	12,085
B1.01D	Day Night	100,000	1,151	2,017	1.75	16,203
B1.01H	Day Night	110,000	1,149	27,399	23.85	124,297
B1.01J	Day Night	120,000	Navigation Process Fails			
BD01.1	Day Only	100,000	2,191	3,813	1.74	32,784
BDR1.4	3 LM/Orbit	40,000	1,949	3,069	1.58	17,325

Table 7

INITIAL ERRORS FOR MODES C AND D

δR =	300,000 ft
δN =	600,000 ft
$\delta \dot{R}$ =	347 ft/sec.
$\delta \dot{T}$ =	695 ft/sec.
$\delta \dot{N}$ =	695 ft/sec.

The initial covariance matrix is, as before, diagonal with diagonal terms equal to the squares of the corresponding initial errors. Under these conditions, the limit of convergence was reached at a down range error of 1,000 miles. After three orbits, Mode C autonomous navigation reduced errors to a level of 20,000 feet and 20 feet per second.

The magnitude of the convergence region for Mode C suggests a dual mode combination of Mode C and Mode A for cases in which accurate navigation must be performed even through initial errors are large. A test of the dual mode performance used the initial errors in Table 7 with a down range error of 1,000 miles. For this test, Mode C autonomous navigation was performed for three orbits after which Mode A was utilized. After four orbits of dual mode navigation, position errors were reduced to a level of 500 feet; after six orbits, position errors were reduced to a 200 foot level.

Mode D

The region of convergence for Mode D is, like Mode C, large and exceeds the applicability of the S parameter. A sequence of runs was performed holding all errors except down range position constant and equal to values in Table 7. The limit of convergence was found at a down range position error of 2,500 nautical miles.

Mode E

The region of convergence for Mode E was found to be 55,000 feet. Decreasing the initial covariance matrix to a level of one fourth the actual initial error levels had a small beneficial effect upon the region of convergence.

Long Term Divergence

A navigation filter which initially converges can show a long term instability[3]. This instability is caused by the round off in a finite word length computer or by an incorrect modeling of phenomena such as the drag and gravity forces. Figures 7a, 7b, 7c shows the sixteen orbit behavior of the three landmark navigation Modes A, B, and E. Each mode employed an observation schedule of three landmarks per orbit. The most accurate mode, the star-landmark mode, exhibited a divergent character after about eight orbits. The onset of divergence began after about twelve orbits for Modes B and E. In each case divergence was initiated by the round off in the computer; these modeling errors were not employed in simulations.

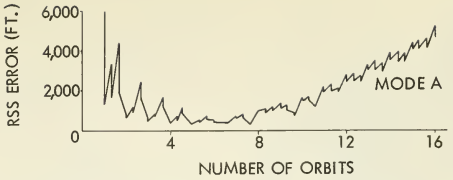


Figure 7a, Sixteen Orbit Behavior of Mode A

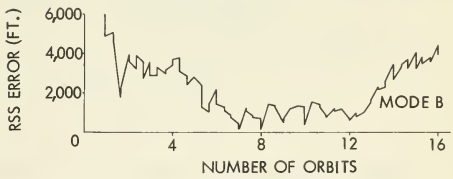


Figure 7b. Sixteen Orbit Behavior of Mode B

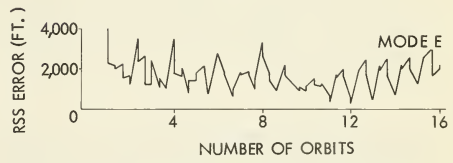


Figure 7c. Sixteen Orbit Behavior of Mode E

Figures 8a and 8b show the long term behavior of Modes C and D. There was no tendency toward long term divergence.

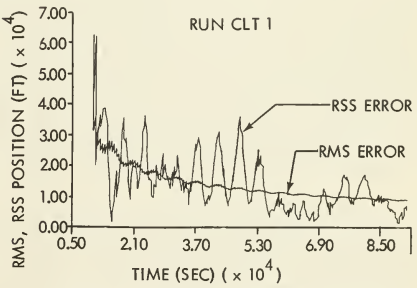


Figure 8a. Sixteen Orbit Behavior of Mode C

Conclusions

Decreasing the magnitude of the initial covariance matrix while holding the magnitude of the actual initial errors constant was found to improve the Kalman filter convergence. On the other hand, increasing the a priori

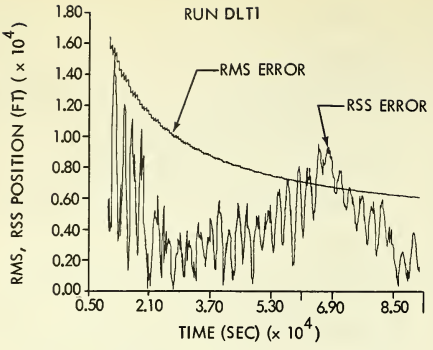


Figure 8b. Sixteen orbit behavior of Mode D

estimate of measurement errors while holding the actual measurement accuracy constant was also found to improve filter convergence. Both variations cause the Kalman filter to place relatively more weight upon the initial orbit estimates and relatively less weight upon measurement data.

For Mode A it was noted that the region of convergence increased as both the actual level of measurement errors and the a priori assumed level of measurement errors increased. The Kalman filter using the less accurate data has a larger region of convergence even though the resulting navigation accuracy is poorer.

Among Modes A, B, and C, Mode A is the most accurate system and has the smallest region of convergence, while Mode C has the largest region of convergence and is the least accurate system. Mode D has a larger region of convergence than Mode C since the added altimeter measurement is a linear function of the state, δR. The region of convergence of the landmark vertical Mode E is larger than that of Mode A, the star landmark mode since the Q matrix is larger in Mode E and the nonlinearity in the measurement is smaller, i.e., the gradient vector is longer for Mode A than Mode E.

Initial estimates of the orbit parameters can be derived from previous navigation data or from a special purpose computation known as a preliminary orbit determination. The region of convergence specifies the required accuracy of the initial estimates. For Modes C and D, the region of convergence is so large that an astronaut could perform preliminary orbit determination with unaided visual sightings through the spacecraft window. For example, the astronaut could estimate his altitude and identify his present latitude and longitude to sufficient accuracy by looking down and locating his approximate ground track relative to a map.

The magnitude of errors in initial estimates derived from previous navigation data depends upon the accuracy of the previous data and the time span during which no new navigation measurements are processed. For example, assuming state-of-the-art injection accuracy, navigation errors will exceed the Mode A convergence region within one half an orbit. The Mode E convergence region will be exceeded within one orbit and Mode B within 2 orbits. Navigation errors will exceed convergence limits for Modes C and D in about six days and fifteen days respectively.

Long term divergence was found to be related to system accuracy. The long term divergence was slightly more pronounced in the more accurate Mode A than in Modes B and E. The least accurate Modes C and D did not show any tendency to long term divergence.

References

1. Bellantoni, J.F. and Dodge, K.W. – "A Square Root Formulation of the Kalman-Schmidt Filter", AIAA paper 67-90, January 1967.

2. Denham, W.F. and Pines, S. – "Sequential Estimation When Measurement Function Nonlinearity is Comparable to Measurement Error", AIAA Journal Vol. 4, No. 6, June 1966.

3. Schlee, F.H., Standish, C.J., and Toda, N.F. – "Divergence in the Kalman Filter", AIAA Journal, Vol. 5, No. 6, pp. 1114-1120, June 1967.

4. Pines, S., Wolf, H., Baile, A., Mohan, J. – "Modifications of the Goddard Minimum Variance Program for the Processing of Real Data," Analytical Mechanics Associates (For Goddard Space Flight Center), October 1966.

5. Toda, N.F. and Schlee, F.H. – "Autonomous Navigation by Optical Tracking of Unknown Landmarks," AIAA/JACC Guidance and Control Conference, Seattle, Washington, August 1966.

Acknowledgment

The authors wish to thank Miss C. Shafer of the IBM Space Systems Center, Endicott, New York, for her significant help in writing the simulation program and obtaining numerical data.

No. 67-624

DRIFT ANGLE ACUITY IN SPACECRAFT ATTITUDE DETERMINATION

by

F. E. NEUMEYER
International Business Machines Corporation
Endicott, New York

AIAA Paper
No. 67-624

AIAA Guidance, Control and Flight Dynamics Conference

HUNTSVILLE, ALABAMA/AUGUST 14-16, 1967

1.03, 7.01

-- NOTES --

DRIFT ANGLE ACUITY IN SPACECRAFT ATTITUDE DETERMINATION

F. E. Neumeyer, Senior Associate Engineer

International Business Machines Corporation
Federal Systems Division, Space Systems Center
Endicott, New York

Summary

The drift angle of a spacecraft is the angle measured on a plane tangent to the earth's surface at the spacecraft's nadir between the spacecraft x-axis and the ground track. This angle can be sensed and measured by image-sensing and computing devices such as the gyrocompassing performed on Gemini and the V/H sensing performed in current military aircraft and spacecraft or by the astronaut directly with an optical sight. This study is an examination of man's ability to determine the drift angle with an optical sight. The results of the study indicate that the astronaut can determine this angle reliably within a few minutes of arc.

Introduction

The horizontal angle between the longitudinal axis of a spacecraft and the track (earth's surface ground path) is known as the yaw angle or the drift angle. The word "drift" is a carry-over from aircraft navigation usage, in that the wind vector and the airspeed-true heading vector resolve to the ground speed-track vector. It is said that the wind "drifts" the aircraft from the heading to the track by an amount known as "drift" angle. The "airspeed-true-heading" vector considered in the aircraft dead reckoning problem is not analogous to the spacecraft problem in extra atmospheric orbit, but is applicable and considerable in the atmospheric flight phase of a lifting reentry vehicle. The ability of an astronaut observer to align a telescope reticle with the apparent direction of motion of imaged surface features may be called drift angle acuity. Errors in drift angle determination are necessarily made up of:

a. System errors
b. Drift angle discrimination error
c. Drift angle reading errors

System errors (a) are the result of inaccuracies in basic system alignment beyond calibration capability. Drift angle discrimination error (b) is the difference between the actual drift and the reticle setting measured in degrees of arc. The extent of this error is the measure of drift angle acuity. Drift angle reading errors (c) are inaccuracies or errors in the visual task of reading a vernier scale, which subjects have been studied by Jacobs[1], Baker[2], Leibowitz[3], and others. The aims of this study are to investigate the influence of various parametric variables on drift angle acuity. The variables considered were the dynamic elements of drift observation geometry, namely the rate of apparent image motion (measured in seconds of time in field of view), the pitch angle of the optical axis of the view finder or driftmeter, and the azimuth of apparent image motion related to the spacecraft axis.

Two tasks are normally required of the astronaut observer: Perception of parallelism (or absence thereof) between the lines of the reticle and the apparent direction of motion of imaged surface features, and alignment of the reticle with this apparent direction of motion so as to produce parallelism. The second task is seen as an iterative process involving the first task and then the manual movement of the reticle azimuth control in the appropriate direction and extent until the threshold of acuity is reached. In this study, perception of parallelism and correction of "error" have both been addressed simultaneously.

Procedure

A. APPARATUS: A Bendix Driftmeter is mounted in a horizontal position above a drum (see Figure 1). The line of sight is directed to the surface of the drum by means of a mirror. Imagery in the form of aerial mosaics is attached to the drum. The scale of the imagery is adjusted to simulate the scale of actual terrain when observed with appropriate magnifications. The imagery is illuminated by two 60 watt light bulbs in reflectors, which afford even illumination without glare as might be encountered when observing the earth from a spacecraft at solar noon. The drum is driven (rotated) by a variable speed motor and transmission capable of duplicating actual image velocities to correspond with image magnification. The drum is rotatable in azimuth on a turntable mount. The turntable incorporated an angle scale and vernier which is directly readable to 1.5 arc minutes. The scale of the driftmeter reticle, also equipped with a vernier, is readable directly to 6 arc minutes. The reticle of the driftmeter is superimposed on the object image in the optical train by rear-illumination and reflection through a prism system. The reticle illumination source is variable by means of a potentiometer available to the subject. The

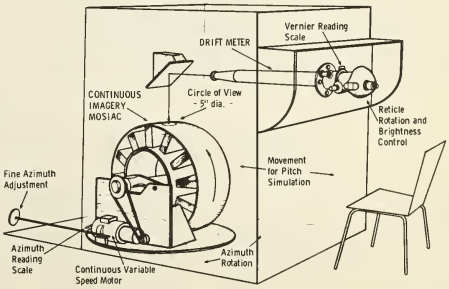


Figure 1. IBM Drift Situation Simulator

67-624

system was aligned with a theodolite and the discrepancy between drum azimuth zero and driftmeter zero introduced into all readings as a calibration correction.

The relationship between image scale (and proportionate image velocity) and the real spacecraft situation is shown in Table 1. It is based on a spacecraft optical telescope with a 48° apparent field-of-view at 1 x magnification.

B. METHODS: The spacecraft image and illumination situation was simulated as close as practicable. The following parameters were assumed:

Spacecraft altitude: 150 nm (900,000 feet)
Ground velocity: 25,000 fps
Optical system magnification: 6x to 90x
Illumination: Simulation of solar noon at nadir

The following variables were considered:

Magnification: 3x to 90x
Pitch angle: 0, 22-1/2°, 45° (forward)
Imagery: a. Photographic mosaics with scales to correspond to magnification: 1/300,000; 1/60,000; 1/30,000; 1/20,000
b. 1/8" Black dots placed at random against a white background.
Image velocity: To correspond to magnification and spacecraft velocity
Subjects: Trained aerial observers

The basic parameter of change in this experiment was assumed to be apparent velocity of images across the field-of-view. For this reason, image velocity (and corresponding image scale) was varied from near-subliminal values to a value at which all images were blurred streaks across the field-of-view. The actual azimuth angle (drift angle) of the imagery was varied about a selected azimuth ±10° so that the subject would not have the advantage of repeating any one unknown angle. The subject was required to rest between sets of observations because of the physical and psychological strain of this unusual effort.

The subjects used were experienced aerial observers familiar with pilotage and dead reckoning instrumentation and problems, psychomotor response to stimuli in this environment, and the use of optical-mechanical apparatus. The subjects had normal uncorrected vision. It was observed that the variability of the subjects' performance decreased with experience. For this reason the subjects were given practice sessions as the data was sought only from a situation of current experience.

The imagery consisted of aerial mosaic photography, scaled appropriately for the orbit and the optical system geometry, and of black dots against a white background.

Image speed settings were used for subliminal values to values above the acuity threshold. Each record session consisted of a number of samples, each sample being a set of observations of drift at one value of image velocity. Data from each sample was recorded and processed to determine sample mean, standard deviation "σ" probable error (0.6745σ), and the 0.95 probability level (1.96σ). The sample means indicate personal bias, which is not dealt with in this study, as various subjects' means showed no distinct regression. The significant statistic is seen as a function of the variance, more meaningful displayed as the "probable error" from the mean. This statistic indicates the inter-quartile range about the mean, within which 0.5 of the variates occur. Also of interest is the commonly used 0.95 probability level. We wish to make a statement such as "the probability that an astronaut can read the drift angle at an image velocity factor of x within a range of ±y' is 0.95". The data is therefore displayed in terms of these probability levels for x as the image velocity factor in seconds of time across the field of view, and y' as the drift angle acuity in minutes of arc from 0 (see Figure 2). It is assumed that sample means can be reduced to 0 by a correction factor.

The acuity range, theoretically infinity at 0 velocity, converges rapidly upon perception of image movement to an apparent minimum between "0" and 3x magnification (or between infinity and 10 seconds time-across-field-of-view. This convergence does not appear on the graph because it is theoretical in nature.

Results

At these low velocities, the image is apparently moving fast enough so that the subject does not forget relative distances between the image and the reticle line as the image progresses, yet not so fast that it is a strain to follow. The phenomenon of perception may not be unrelated to vernier acuity. The acuity curves are seen to diverge from these minima. At the higher velocities the image is apparently traveling too fast to follow or to perceive the distance between the image

Table 1

RELATIONSHIP BETWEEN IMAGE SCALE AND REAL SPACECRAFT SITUATION

Magnification	Real Field	Diameter		Scale: 1	Image Velocity Factor (Time Across Field of View)
		N.M.	Feet		
6x	8°00'	20.94	126,500	300,000	5.0 sec
30x	1°36'	4.19	25,300	60,000	1.0 sec
60x	48'	2.09	12,650	30,000	0.5 sec
90x	32'	1.40	8,430	20,000	0.33 sec

and the line at successive image points. Some improvement was expected with greater velocities as images "streaked" into almost recognizable lines, but none was noted.

A scatter diagram plot of the data in Figure 2 seemed to indicate curvilinear regression so the data was fitted to a second degree equation by the method of least squares. The plot resolved as linear, however, as the third term of the quadratic was negligible. The coefficient of correlation (degree of linear dependence that y has on x) for this data is +0.76, where

$$r = \pm\sqrt{1 - \frac{\text{Sum of Squares about regression}}{\text{Sum of Squares total}}}$$

Note that (in Figure 2, 3, 4, 5) acuity is ± the noted x-axis scalar values. Only the + values are graphed.

A number of samples were taken at 30x magnification (in which an image is in the field of view for 1 second) introducing image scale as the dependent variable. No significant difference in acuity was noted, which introduces the supposition that the eye resolves and the brain interprets images as appropriate to the problem without regard to actual image size or relative magnification. Changes in reticle pattern were also of negligible effect.

It may be tentatively concluded from this that the subjects attention and concentration is focused upon details of the imagery rather than on any conception of scalar relationships.

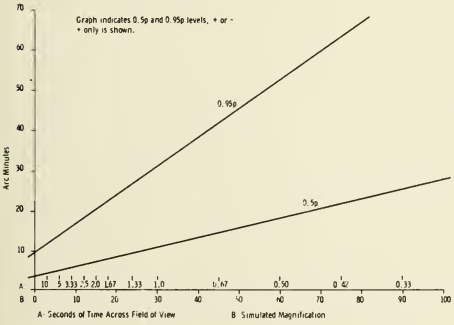


Figure 2. Drift Angle Acuity (Error Correction) Using Photo Imagery (All Subjects)

Figure 3 and 4 portray results obtained when the pitch angle of view was changed from 0° to 22 1/2° forward of the vertical and 45° respectively. The curves were sketched from the scatter diagrams: their curvilinear shape indicates a second-degree equation but was not fitted as such. These plots correlate with the data in Figure 2 by coefficients of +0.89 and +0.44 respectively. (The correlation coefficient varies from -1.0 to +1.0). The latter figure (+0.44) indicates that either insufficient data was considered or that some basic cause exists for lack of positive correlation.

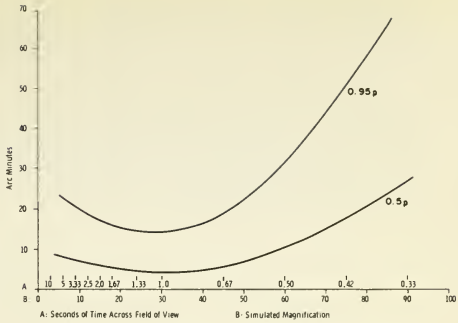


Figure 3. Drift Angle Acuity (Error Correction) Using Photo Imagery Azimuth ϕ^{o} Pitch 22 1/2°

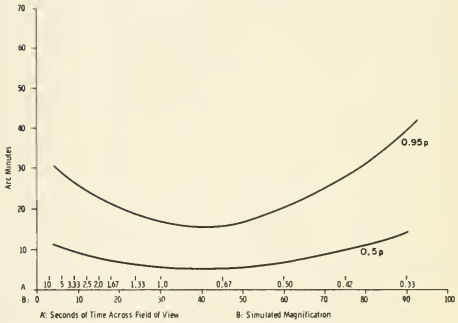


Figure 4. Drift Angle Acuity (Error Correction) Using Photo Imagery Azimuth ϕ^{o} Pitch 45°

The final graph, Figure 5, displays results similar to Figure 2 but for data that was acquired from the observation of artificial imagery, i.e., 1/8" black dots scattered at random on a white background. Acuity improves, as could be expected, by use of this maximum-contrast stimuli. Both types of imagery

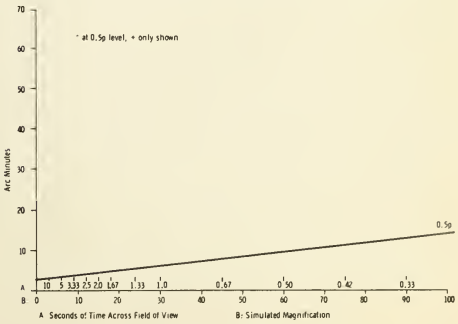


Figure 5. Drift Angle Acuity (Error Correction) Using Artificial Imagery (All Subjects)

were examined, as the psychologist on one hand would perfer the close-to-absolute nature of the artificial imagery stimuli, while the astronaut might prefer data more closely related to the real image situation.

The evidence points to the conclusion that the astronaut can read the drift angle in the 0 to 20x magnification range (or more precisely in the infinity-to-1.5 seconds-in-the-field-of-view range) with an acuity from 3 to 8 minutes of arc or better half of the time, or from 10 to 25 minutes of arc at these velocities 95% of the time.

References

1) D. H. Jaclobs "Fundamentals of Optical Engineering", 1943, pp. 86-88.

2) K. E. Baker "Journal of the Optical Society of America", 39, 1949, pp. 567-576.

3) H. Leibowitz "American Journal of Psychology", June 1955, Volume 68, No. 2 pp. 266-273.

-- NOTES --

-- NOTES --