

MODULAR ELECTRONICS LEARNING (MODEL) PROJECT



RADIO ANTENNAS

© 2020-2025 BY TONY R. KUPHALDT – UNDER THE TERMS AND CONDITIONS OF THE
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL PUBLIC LICENSE

LAST UPDATE = 22 JULY 2025

This is a copyrighted work, but licensed under the Creative Commons Attribution 4.0 International Public License. A copy of this license is found in the last Appendix of this document. Alternatively, you may visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons: 171 Second Street, Suite 300, San Francisco, California, 94105, USA. The terms and conditions of this license allow for free copying, distribution, and/or modification of all licensed works by the general public.

Contents

1	Introduction	3
1.1	Recommendations for students	3
1.2	Challenging concepts related to antennas	5
1.3	Recommendations for instructors	6
2	Case Tutorial	7
2.1	Example: making a 900 MHz antenna	8
2.2	Example: cable loss decibel calculations	15
3	Tutorial	17
3.1	Basic antenna types	17
3.2	Antenna size	19
3.3	Dipole voltage and current distributions	22
3.4	Antenna polarization and directionality	24
3.5	Antenna reciprocity	28
3.6	Antenna characteristics	29
3.6.1	Vertical monopole antenna	30
3.6.2	Vertical dipole antenna	31
3.6.3	Vertical loop antenna	32
3.6.4	Vertical Yagi antenna	33
3.6.5	Vertical corner-reflector antenna	36
3.6.6	Vertical log-periodic dipole antenna	37
3.6.7	Circularly-polarized antennas	38
3.7	Antenna gain calculations	40
3.8	Antenna standing wave ratio	43
4	Derivations and Technical References	45
4.1	Decibels	46
4.2	Near-field versus far-field regions	56
4.3	The Poynting vector	60
4.4	Maxwell's equations of electromagnetism	64
4.4.1	Gauss's Law of electric fields	64
4.4.2	Gauss's Law of magnetic fields	64
4.4.3	Faraday's Law of electromagnetic induction, extended	65

CONTENTS	1
4.4.4 Ampère's Circuital Law, extended	65
4.4.5 Proof of electromagnetic waves	66
5 Questions	69
5.1 Conceptual reasoning	73
5.1.1 Reading outline and reflections	74
5.1.2 Foundational concepts	75
5.1.3 Parabolic dish feed	77
5.1.4 Antenna blind spots	78
5.1.5 Dummy antenna	78
5.1.6 Microwave detector	79
5.1.7 Transmitter testing	80
5.1.8 Close-coupled antennas	80
5.2 Quantitative reasoning	81
5.2.1 Miscellaneous physical constants	82
5.2.2 Introduction to spreadsheets	83
5.2.3 Antenna dimensions	86
5.2.4 Transmitter power and field strength	87
5.2.5 Monopole radiation strength	89
5.3 Diagnostic reasoning	91
5.3.1 Whip antenna orientations	91
5.3.2 Incorrect decibel calculation program	92
A Problem-Solving Strategies	93
B Instructional philosophy	95
B.1 First principles of learning	96
B.2 Proven strategies for instructors	97
B.3 Proven strategies for students	99
B.4 Design of these learning modules	100
C Tools used	103
D Creative Commons License	107
E References	115
F Version history	117
Index	119

Chapter 1

Introduction

1.1 Recommendations for students

Antennas are deceptively complex components in an RF (“Radio Frequency”) system, because at first inspection they resemble nothing more than some pieces of wire (or even a *single* piece of wire). However, the parasitic capacitance and inductance inherent to those pieces of wire make antennas complex indeed.

This module is a very shallow introduction to the very deep world of antenna function and design. The fact that RF engineers may easily specialize in their careers as antenna designers gives you some indication of just how complex this subject is. In fact, antenna design is complex enough that computers running genetic algorithms have been used to *evolve* novel antenna designs by simulation-testing random shapes until an optimum shape emerges from the trials!

Important concepts related to the topic of antennas include **frequency** versus **wavelength**, **electromagnetic radiation**, **resonance**, **fundamental** versus **harmonic** frequencies, **polarization**, **reciprocity**, **directionality**, **antenna gain**, **dipole antennas**, **monopole antennas**, **loop antennas**, **transmission lines**, **standing wave ratio**, signal **reflections**, **radiation patterns**, **SCADA** systems, **RTU** and **MTU** units, **radiation resistance**, **balanced** versus **unbalanced** signals, **decibels**, and the **Law of Energy Conservation**.

Here are some good questions to ask of yourself while studying this subject:

- How might an experiment be designed and conducted to explore the concept of antenna polarization? What hypothesis (i.e. prediction) might you pose for that experiment, and what result(s) would either support or disprove that hypothesis?
- How might an experiment be designed and conducted to measure the gain of an antenna? What hypothesis (i.e. prediction) might you pose for that experiment, and what result(s) would either support or disprove that hypothesis?
- How might an experiment be designed and conducted to map the radiation pattern of an antenna? What hypothesis (i.e. prediction) might you pose for that experiment, and what result(s) would either support or disprove that hypothesis?

- How does an antenna work to radiate (i.e. transmit) electromagnetic waves?
- How does an antenna work to intercept (i.e. receive) electromagnetic waves?
- How does the physical size of an antenna relate to its ideal operating frequency?
- Why is the physical orientation of an antenna important to its function?
- How can an antenna have a *gain* as a purely passive device, without violating the Law of Energy Conservation?
- What purpose is served by the “reflector” and “director” elements of a Yagi antenna?
- What useful information may be gleaned from an antenna’s radiation pattern diagram?
- Must transmitting and receiving antennas be differently designed, or will the same design work equally well in both modes?
- What is an isotropic antenna, and why is the concept an important one?
- Is an antenna an electrical load, source, or something else?
- How can you tell if an antenna’s length has been properly trimmed for the transmitter it connects to?
- Where might a directional antenna be best applied?
- What is a *balanced* signal, and how does this differ from an *unbalanced* signal?
- What do the following terms mean: *dB*, *dBd*, and *dBi*?
- How does a circularly-polarized antenna work?

1.2 Challenging concepts related to antennas

The following list cites concepts related to this module's topic that are easily misunderstood, along with suggestions for properly understanding them:

- **Parasitic effects** – phenomena such as resistance, capacitance, and inductance don't just exist where we want them to, but in fact always exist everywhere they can. All conductors (except for superconductors) have some amount of resistance, all conductors separated by an insulating gap exhibit capacitance, and all conductors possess inductance by virtue of their ability to produce a magnetic field with the passage of current. Often these "parasitic" effects are small enough to ignore, but in some applications such as radio circuitry their presence is in fact absolutely necessary for the system to work at all!
- **Decibels** – "decibels" are an attempt to express power ratios (i.e. power gains or attenuation factors) logarithmically rather than linearly, and as such they tend to generate confusion for students less familiar (or unfamiliar) with exponential and logarithmic functions.
- **Antenna gain** – gain for any antenna is most properly calculated as the ratio of its transmitting/receiving capability in a given direction compared to that of some reference antenna, usually a theoretical isotropic antenna or in some cases a simple dipole antenna. Unlike amplifier gain where signal power actually gets boosted, antenna gain is a kind of pseudo-gain and really just expresses the degree to which a given antenna structure is more directional than the reference antenna.
- **Maxwell's equations** – the existence of electric and magnetic fields within empty space, with no conductors nearby, was a discovery made by James Clerk Maxwell in the late 1800's, and is as nonintuitive today as it was then. RF systems with their extremely high frequencies manifest effects related to detached electric and magnetic fields that are all but undetectable in lower-frequency systems.

1.3 Recommendations for instructors

This section lists realistic student learning outcomes supported by the content of the module as well as suggested means of assessing (measuring) student learning. The outcomes state what learners should be able to do, and the assessments are specific challenges to prove students have learned.

- **Outcome** – Demonstrate effective technical reading and writing

Assessment – Students present their outlines of this module’s instructional chapters (e.g. Case Tutorial, Tutorial, Historical References, etc.) ideally as an entry to a larger Journal document chronicling their learning. These outlines should exhibit good-faith effort at summarizing major concepts explained in the text.

- **Outcome** – Apply the concept of radiation patterns to specific antenna applications

Assessment – Identify where a given type of antenna is “blind” to incoming signals; e.g. pose problems in the form of the “Antenna blind spots” Conceptual Reasoning question.

Assessment – Identify signal attenuation at different angles to a transmitting antenna given a decibel-scaled radiation pattern for that antenna; e.g. pose problems in the form of the “Monopole radiation strength” Quantitative Reasoning question.

- **Outcome** – Independent research

Assessment – Read and summarize in your own words reliable historical documents on the development of radio technology.

Assessment – Locate antenna datasheets and properly interpret some of the information contained in those documents including radiation resistance, bandwidth, radiation pattern, physical dimensions, power rating, gain, etc.

Chapter 2

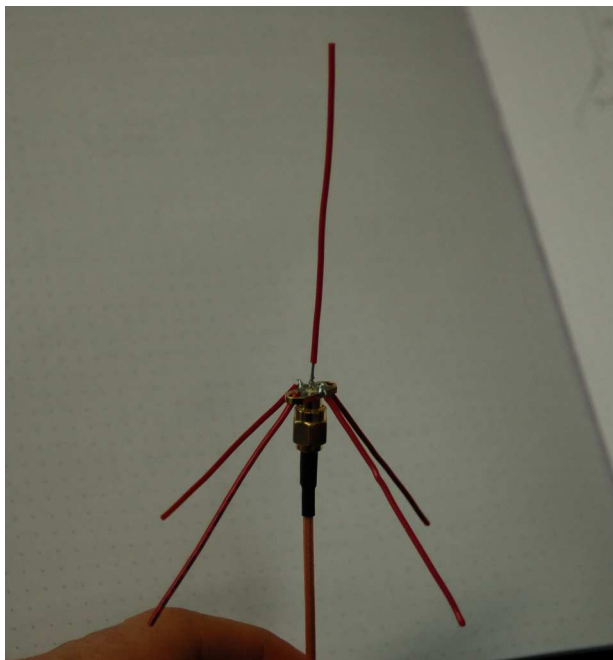
Case Tutorial

The idea behind a *Case Tutorial* is to explore new concepts by way of example. In this chapter you will read less presentation of theory compared to other Tutorial chapters, but by close observation and comparison of the given examples be able to discern patterns and principles much the same way as a scientific experimenter. Hopefully you will find these cases illuminating, and a good supplement to text-based tutorials.

These examples also serve well as challenges following your reading of the other Tutorial(s) in this module – can you explain *why* the circuits behave as they do?

2.1 Example: making a 900 MHz antenna

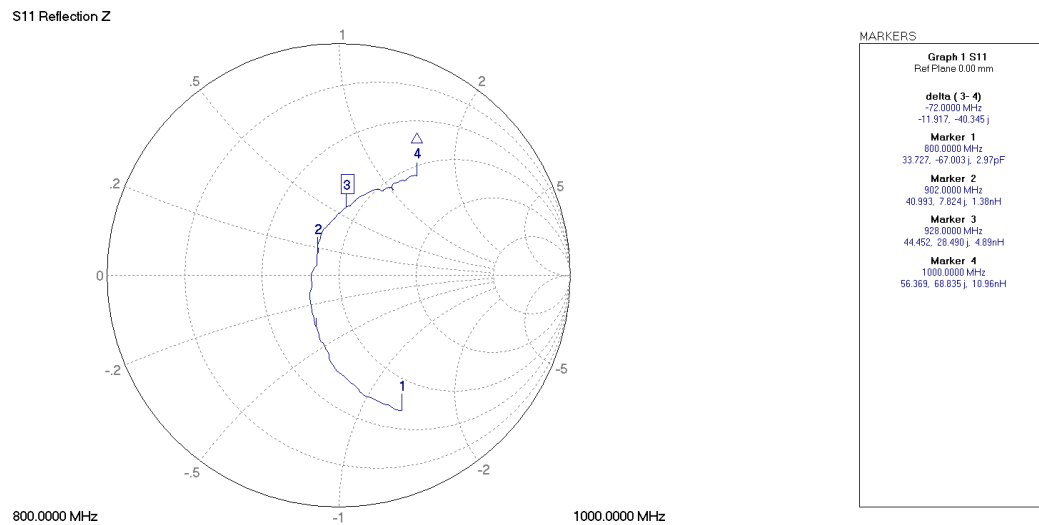
A great many industrial and consumer grade electronic devices use the 902 MHz to 928 MHz frequency band which is one of several “ISM” (Industrial, Scientific, Medical) bands requiring no FCC license within the United States if operating below specified maximum EIRP power levels. Making your own monopole (“whip”) antenna for one of these devices is easy to do using an RF connector and a set of solid metal wires for the driven monopole element and the radial counterpoise elements (used to form a “ground plane”) for the driven element, as we see in the following photograph:



When the radio transceiver uses a coaxial connector for the antenna connection as is the case in our application, the signal is necessarily unbalanced. Since a monopole antenna is also unbalanced, there will be no need for a *balun* between the two as would be the case if we used a balanced antenna design such as a dipole.

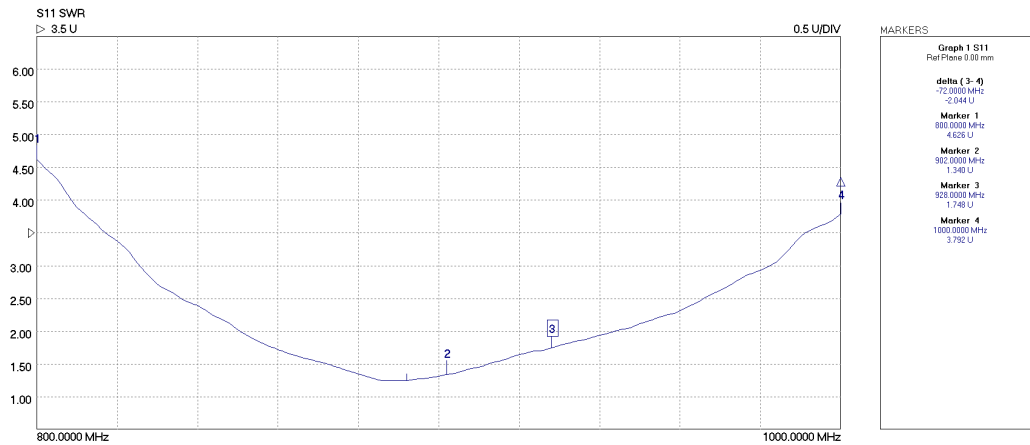
The wavelength of a 902 MHz signal is approximately 0.3324 meters, with one-quarter of that being 83.1 millimeters. This will be the (ideal) length of the driven monopole element. Radial lengths are less critical, and in this case were 72 millimeters each. By using a suitable test instrument such as a vector network analyzer (VNA) or impedance analyzer, we will be able to measure the antenna to determine its fundamental resonant frequency, and to trim the driven element wire length to obtain the best results.

When tested on a VNA sweeping between 800 MHz and 1 GHz the following result was obtained and plotted in the form of a Smith Chart showing the antenna's impedance as a complex-numbered value:



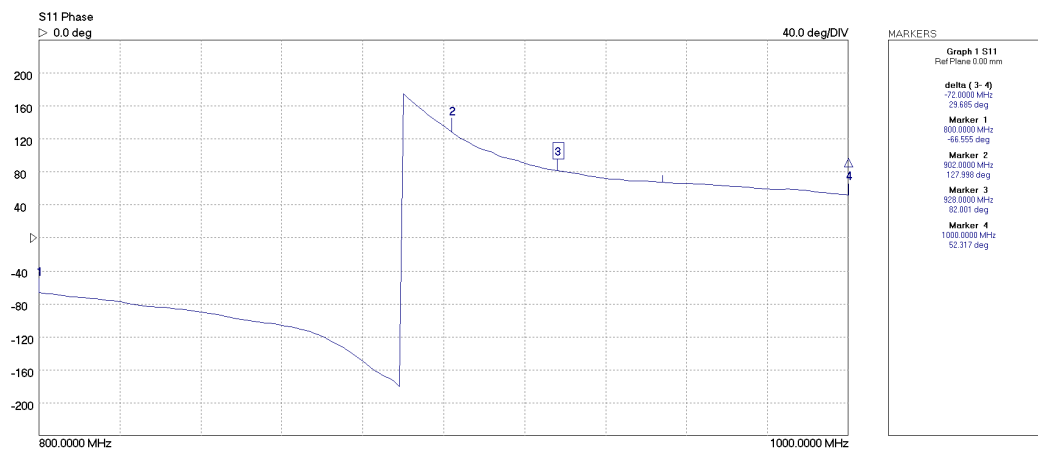
Markers 2 and 3 represent the bounds of our ISM band, 902 MHz and 928 MHz respectively. We achieve a non-reactive (i.e. purely resistive) impedance where the trace crosses the horizontal line inside the Smith chart, and this appears to happen at a frequency *below* 902 MHz. This means our 83.1 millimeter driven element length is in reality slightly too long for the 902-928 MHz frequency range. The reason for this is a phenomenon known as *end effect* where the effective capacitance at the tip of the driven element is greater than what it would be if the wire were infinite in length (i.e. no end). This extra capacitance results in any antenna's driven element resonating at a frequency that is *lower* than what we ideally predict based on signal wavelength.

Another useful plot of antenna performance is its *Standing Wave Ratio*, or SWR. The ideal and lowest-possible SWR value for any antenna is unity (1), with any value greater than that being less than ideal:



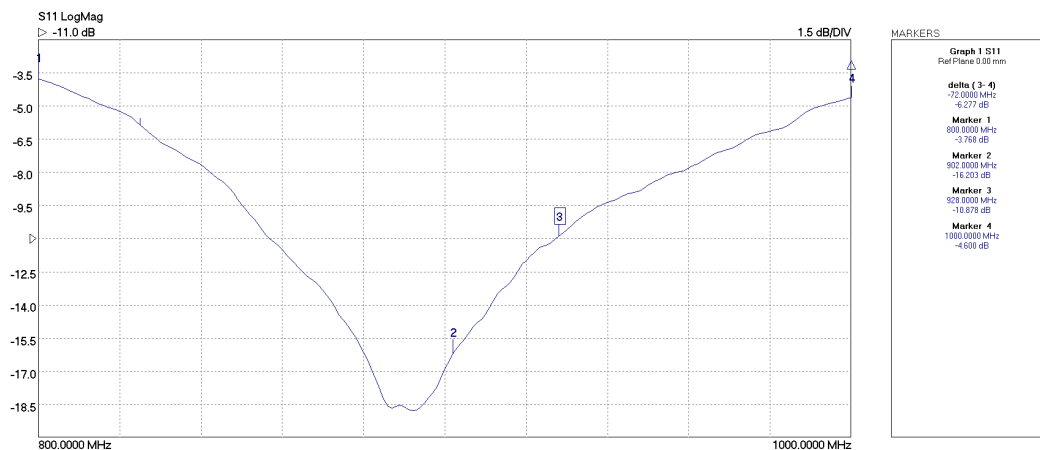
Here we see the SWR achieving its lowest value at a frequency that is one again below 902 MHz. This also tells us the antenna's driven element is physically too long.

Perhaps the most dramatic plot showing the antenna's resonant frequency is the VNA's *phase* measurement, showing the polar phase angle of the antenna's measured impedance. At resonance that phase angle should be zero. Immediately above or below resonance the impedance phase angle will jump to a large positive value (inductive) or negative value (capacitive), respectively. The sharp vertical “jump” we see in this phase plot represents the narrow frequency over which the antenna transitions from predominantly capacitive to predominantly inductive:



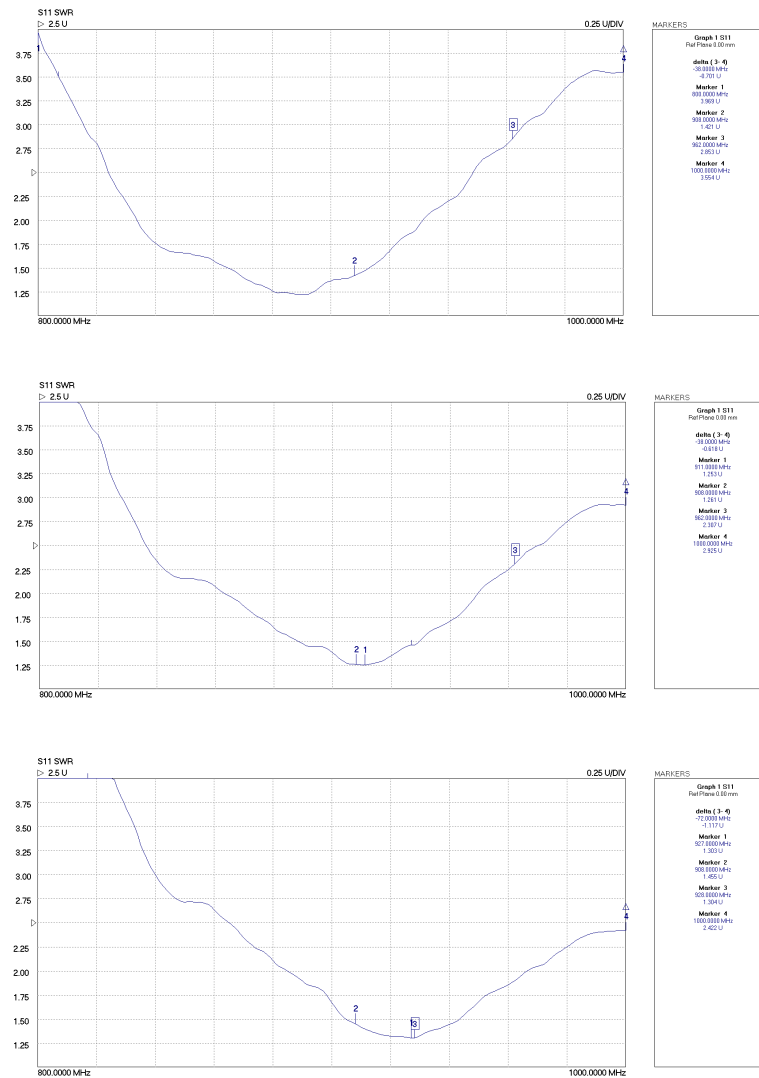
Once again this transition occurs below 902 MHz (to the left of marker #2), telling us the antenna's driven monopole element is physically too long.

Lastly, we see the VNA's plot showing *reflected signal magnitude* in decibels. Ideally an antenna radiates all energy received from the wired signal source into space, reflecting none of that energy back to the signal source. So, ideally the reflected signal value at the antenna's operating frequency should be zero in magnitude, or infinitely less than the "forward" energy sent to the antenna from the source. A large negative decibel value, therefore, is what we should expect to see at the operating frequency for the reflected signal magnitude:



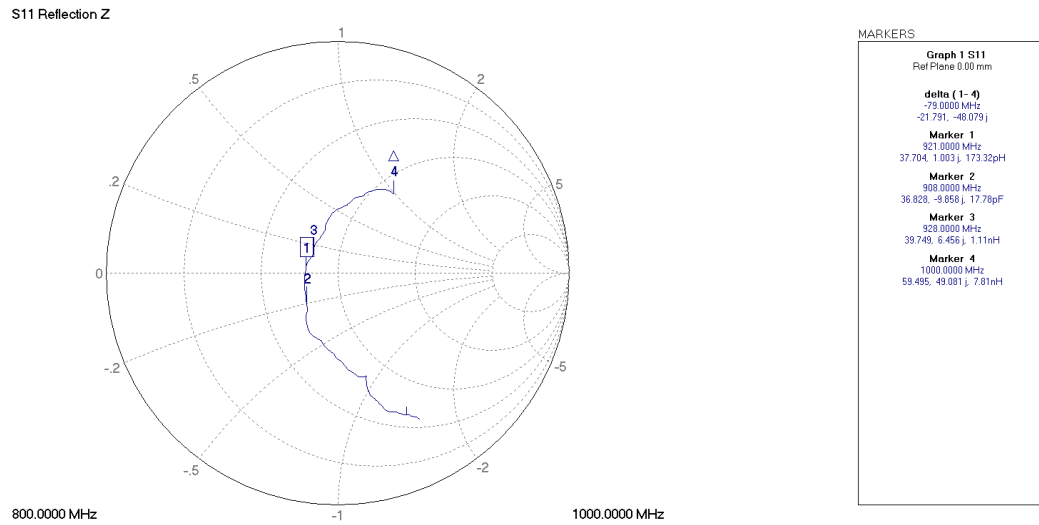
As with the other measurements, we see this ideal point occurring at a frequency below marker #2 (902 MHz) which tells us the driven element is physically too long and needs to be trimmed shorter.

In these next three screenshots, we see SWR measurements taken of this antenna with the original length (top), trimmed by 3 millimeters (middle), and trimmed an additional 2.25 millimeters (bottom):



After the first 3 millimeter trimming the SWR reaches its lowest value at a frequency of 911 MHz (marker #1). After the additional 2.25 millimeter trimming the SWR minimum point is at 927 MHz, again shown by marker #1. As you can see, the process of trimming an antenna at this frequency range is a delicate one indeed!

With the SWR minimum at 927 MHz, the Smith chart showed an antenna impedance of $37.7 + j1$ Ohms at that frequency which is nearly perfectly resistive:



2.2 Example: cable loss decibel calculations

If a radio transmitter sends 15 Watts of power into a coaxial cable having a loss rating of -0.16 dB per meter, and the cable is 20 meters in length, the power available at the far end of the cable must be **7.1795 Watts**.

If an RF signal source sends 10 mW into one end of a 5-meter cable and the other end of the cable outputs 8 mW of signal power, the cable must have a loss rating of **-0.194 dB per meter**.

An RF signal of 2 Watts exits a cable 3 meters in length having a loss rating of -0.14 dB per meter. The amount of power being input to this cable must be **2.203 Watts**.

An RF signal of 15 Watts enters one end of a cable having a loss rating of -0.21 dB per foot, and an RF signal of 10 Watts exits the other end of that same cable. The cable's length must be **8.38530 feet**.

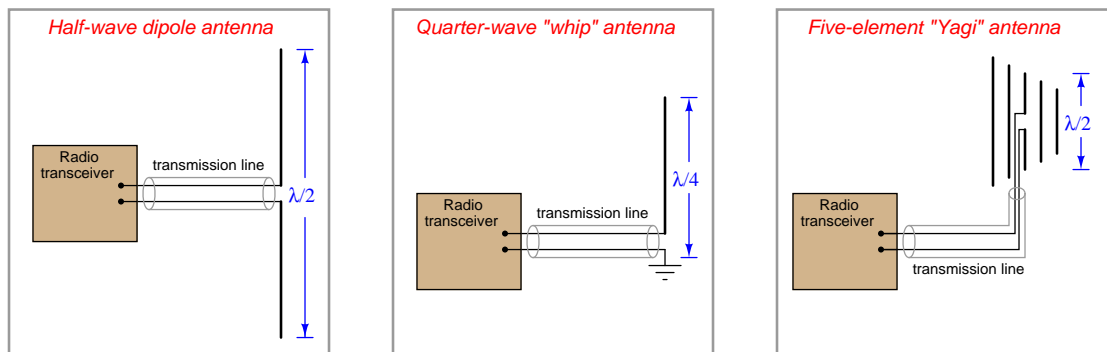
A 50-Watt radio transmitter connected to a 12-foot cable having a loss rating of 0.041 dB/foot terminates at a lightning arrestor with a loss rating of 0.5 dB, and finally to an antenna. The amount of power received by this antenna will be **39.7896 Watts**.

Chapter 3

Tutorial

3.1 Basic antenna types

A *radio wave* is a form of electromagnetic radiation, comprised of oscillating electric and magnetic fields, with the *wavelength* being the physical distance traveled by the wave in the time of one cycle (period). An *antenna* is nothing more than a conductive structure designed to emit radio waves when energized by a radio-frequency (“RF”) electrical power source, and/or generate radio-frequency electrical signals when intercepting radio waves. Three common antenna designs appear here:



Antennas don’t begin to behave as emitters or collectors of RF energy unless and until their physical dimensions approach a fraction of the signal’s wavelength.

The Yagi antenna, with its “director” and “reflector” elements fore and aft of the dipole element, exhibits a high degree of *directionality*, whereas the dipole and monopole (“whip”) antennas tend to emit and receive electromagnetic waves equally well in all directions perpendicular to their axes. Directional antennas are ideal for applications such as radar, and also in point-to-point communication applications. Omnidirectional antennas such as the dipole and whip are better suited for applications requiring equal sensitivity in multiple directions.

A photograph of an actual Yagi antenna used in a SCADA system appears here:



The wavelength (λ) of any wave is its propagation velocity divided by its frequency. For radio waves, the propagation velocity is the speed of light (2.99792×10^8 meters per second, commonly represented as c), and the frequency is expressed in Hertz:

$$\lambda = \frac{c}{f}$$

Antenna dimensions are related to signal wavelength because antennas work most effectively in a condition of electrical *resonance*, and an antenna's geometry dictates the amount of parasitic capacitance and inductance its elements possess. In other words, the physical size of the antenna is such that it will electrically resonate at certain frequencies: a *fundamental* frequency as well as certain *harmonics* (integer-multiples) of that fundamental frequency. Non-resonant antenna designs exist as well, but even with these there is a definite correlation between the wavelength of the RF signal and the physical dimensions of the antenna. For this reason, antenna size is inversely proportional to signal frequency: low-frequency antennas must be large, while high-frequency antennas may be small.

For example, a quarter-wave “whip” antenna designed for a 900 MHz industrial transceiver application will be approximately¹ 8.3 centimeters in length. The same antenna design applied to an AM broadcast radio transmitter operating at 550 kHz would be approximately *136 meters* in length!

¹Due to the “end effect” of lumped capacitance at the tip of the antenna, an actual quarter-wave antenna needs to be slightly shorter than an actual quarter of the wavelength. This holds true for dipoles and other antenna designs as well.

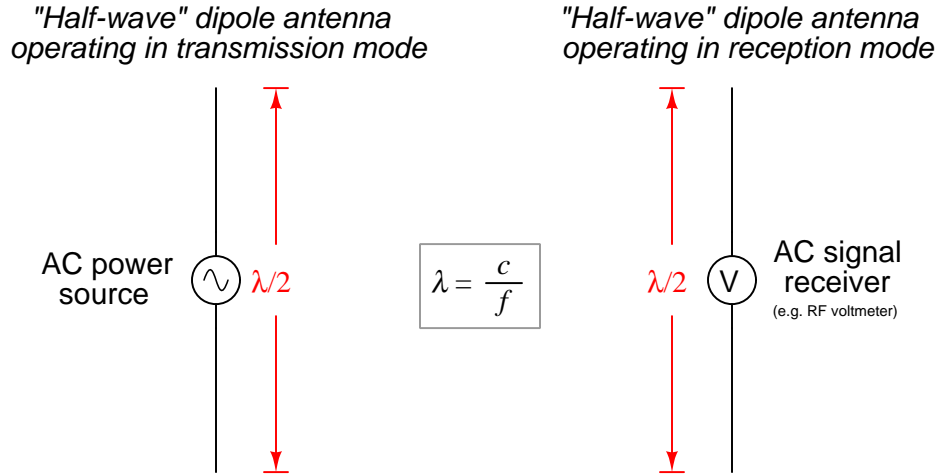
The following photograph shows a *half-wave* “whip” antenna, located at the roofline of a building. The additional length of this design makes it more efficient than its quarter-wave cousin. This particular antenna stands approximately one meter in length from connector to tip, yielding a full wavelength value (λ) of 2 meters, equivalent to 150 MHz:



3.2 Antenna size

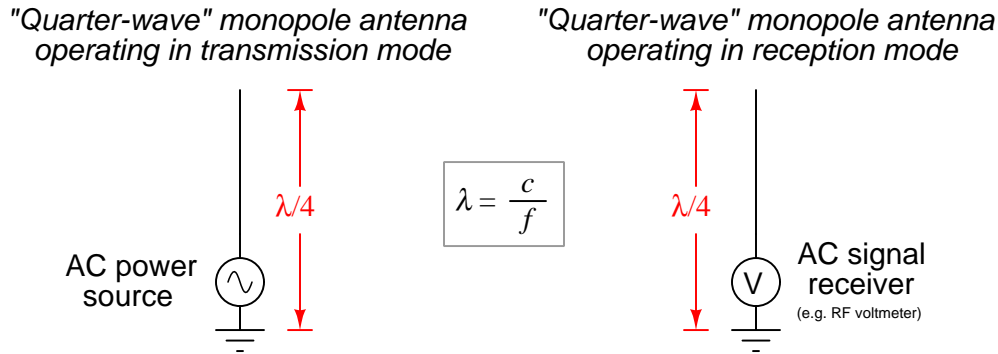
Antennas are fundamentally *resonant* components: their parasitic capacitance and inductance able to store and release energy in the forms of electric and magnetic fields, respectively, and this enables antennas to exchange energy back and forth between these complementary forms. An antenna will function most efficiently – whether transmitting or receiving – when operated in a condition of resonance where the signal frequency matches that of the antenna’s natural oscillations. The relationship between ideal frequency and antenna length is inverse: the longer the antenna, the lower its fundamental frequency, and vice-versa.

Dipole antennas operate best at a wavelength twice as long as the total length of wire:



A “half-wave” dipole antenna with a total length of 5 meters² will radiate optimally at a frequency of 30 MHz, given that the velocity of electromagnetic waves in open space is approximately 3×10^8 meters per second. This same antenna will also effectively resonate at any *harmonic* (integer multiple) of 30 MHz (e.g. 60 MHz, 90 MHz, 120 MHz, etc.) just as a transmission line is able to resonate at a fundamental frequency as well as any harmonic thereof.

A popular variation on the theme of the half-wave dipole antenna is the so-called *monopole*, *quarter-wave*, or “whip” antenna, which is exactly what you might think it is: one-half of a half-wave antenna. Instead of two wires pointed away from each other, we substitute an earth-ground connection for one of the wires:

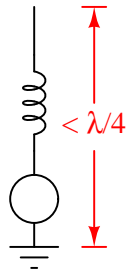


Quarter-wave antennas tend to be less effective than half-wave antennas, but are usually much more convenient to construct for real applications.

²In practice, the ideal length of a dipole antenna turns out to be just a bit shorter than theoretical, due to lumped-capacitive effects at the wire ends. Thus, a resonant 30 MHz half-wave dipole antenna should actually be about 4.75 meters in length rather than exactly 5 meters in length.

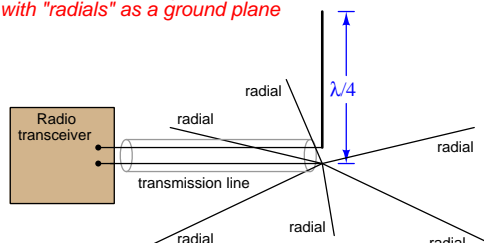
For applications where the ideal length of an antenna is impractically long, resonance may be achieved using a shorter antenna by adding a *loading coil*, usually located in the center of the antenna. The loading coil adds inductance to the antenna structure, reducing the resonant frequency just as increased inductance will decrease resonant frequency for any LC “tank circuit” network. The photograph on the right-hand side shows a Citizen’s Band (CB) antenna magnetically mounted to the roof of a vehicle, enabling the antenna to resonate while being much shorter than the 2.73-2.78 meters normally required for a quarter-wave antenna operating between 26.965 MHz and 27.405 MHz (the “11 meter” band permitted for CB communications in the United States):

"Quarter-wave" monopole antenna with loading coil to reduce length



It is important to understand that quarter-wave (or “monopole”) antennas absolutely require a conductive surface perpendicular to the antenna in order to properly function. In other words, *this conductive plane is really the other half of the antenna*, serving the same purpose as the second element of a dipole. In many installations the Earth itself serves as this conductive surface, and good antenna performance requires an electrically sound connection with the Earth. If the soil beneath an antenna is too dry and therefore not very conductive, it may be augmented by an array of horizontal or angles wires extending radially from the base of the antenna functioning as a *ground plane*, or *counterpoise*.

Quarter-wave "whip" antenna with "radials" as a ground plane

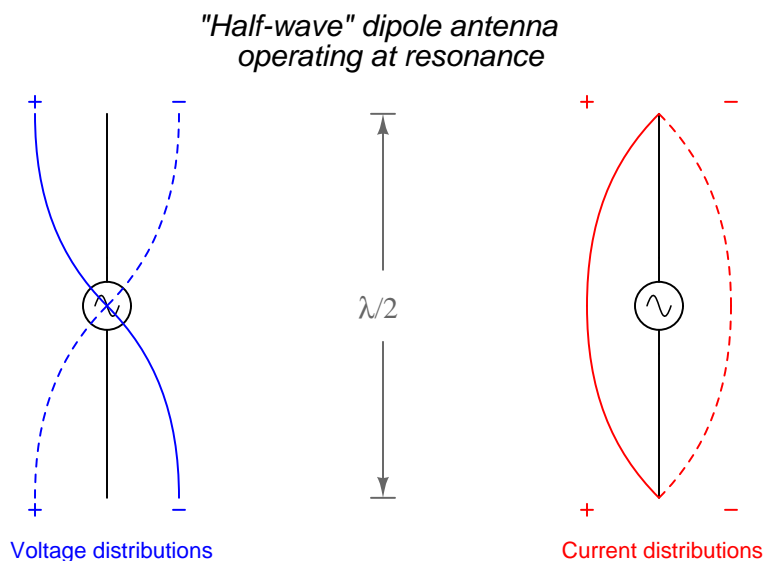


Locating a quarter-wave antenna above a metal-roofed building is another way to achieve a

good ground plane for effective operation. On boats and ships the water itself acts as an excellent ground plane for the antenna, but once again a good “solid” electrical connection between the radio equipment and the water must be established for good performance. For metal-hulled vessels this task is easy: simply use the metal hull itself as the connection point with the water. On fiberglass-hulled vessels some metallic conductor must be inserted into the water to make this connection.

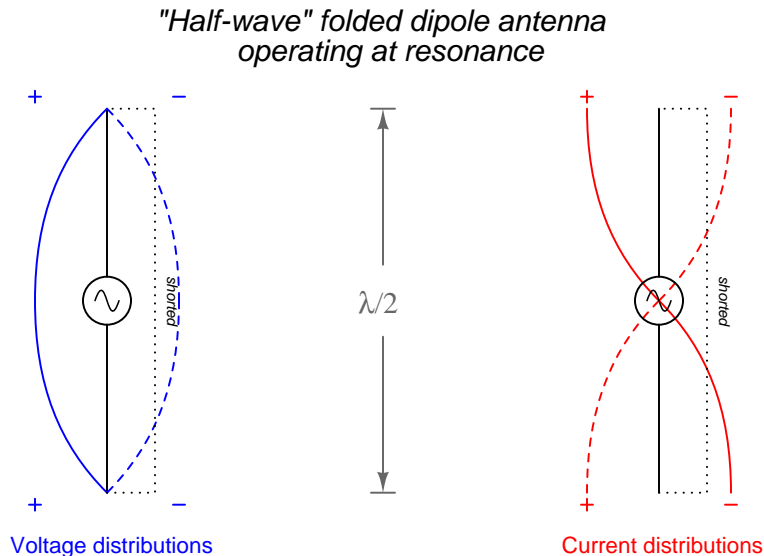
3.3 Dipole voltage and current distributions

Antennas work with signals oscillating so fast that the electric and magnetic fields literally cannot travel (at the speed of light) fast enough to traverse the length of the antenna before a substantial portion of the cycle evolves. This means at any given instant in time there will be different quantities of electrical potential and of electric current at different locations along the metal rods or wires of an antenna’s elements, just as there is for any *transmission line*. Just like transmission lines, antennas may operate in conditions where *standing waves* of voltage and or current form on their conductors. For a classic dipole antenna, which may be thought of as an unshielded transmission line terminated by an “open” at the ends, we find a voltage standing wave with anti-nodes at the tips and a node at the center, and we also find a current standing wave with an anti-node at the center and nodes at the two tips:



The guaranteed consequence of any electrical “open” is the absence of current, and so that is what we see at the tips of the dipole antenna where the metal rods or wires come to their ends. Voltage, however, is perfectly possible at an electrical “open” and so the open tips of the dipole are where we see the peaks (“anti-nodes”) of the voltage standing wave. Note how the AC source “sees” a condition of low voltage and high current, which by Ohm’s Law must mean a low antenna impedance ($Z = \frac{V}{I}$). Typical impedance for a folded half-wave dipole antenna operating at resonance is near 73 Ohms.

What happens if we model an antenna on a transmission line having its far end *shorted* rather than open? In that case we have a *folded dipole* or *loop* antenna. It's difficult to draw convincing waveshapes showing standing voltage and current waves when the antenna is no longer drawn as a straight line, and so we will “cheat” by simply denoting that the ends of the folded dipole are electrically shorted together using a dotted pathway assumed to have zero length:



The guaranteed consequence of any electrical “short” is the absence of voltage, and so that is what we see at the shorted “ends” of the folded dipole antenna where the metal rods or wires come together to form a loop. Current, however, is perfectly possible at an electrical “short” and so the shorted ends of the folded dipole are where we see the peaks (“anti-nodes”) of the current standing wave. Note how the AC source “sees” a condition of high voltage and low current, which by Ohm’s Law must mean a high antenna impedance ($Z = \frac{V}{I}$). Typical impedance for a folded half-wave dipole antenna operating at resonance is near 300 Ohms.

Note the apparent inversion of termination impedance for a half-wave antenna: if the ends (tips) are open then the source sees a low impedance, but if the ends are actually shorted together to form a loop then the source sees a high impedance! Such is the consequence of the AC frequency being so high that the electric and magnetic fields cannot reach the end(s) of the antenna from the center-located source until a quarter of the wave’s cycle has elapsed. While the source sees a peak (anti-node) the ends see a node, and vice-versa.

This is often a point of confusion for students just beginning to study antennas and transmission lines alike: how can such a deceptively simple structure have such non-intuitive electrical properties? The reason behind this confusion is a tendency to analyze the antenna or transmission line in terms of DC. To DC, a simple dipole antenna is completely open and therefore should have infinite impedance, and a folded dipole (loop) is completely shorted and therefore should have (nearly) zero impedance. However, when approaching antennas and transmission lines from the perspective of high-frequency AC where electrical conditions at the far end of the antenna aren’t identical to those at the feedpoint

where the source connects due to the finite speed at which electric and magnetic fields move, we see that what appears to be shorted (to DC) will appear to have a high impedance to AC and vice-versa.

3.4 Antenna polarization and directionality

An important factor affecting radio system performance is the orientation of the receiving antenna with respect to the transmitting antenna. Antenna orientation matters because electromagnetic waves are *polarized*: their electric and magnetic fields oscillating in specific spatial planes (at right angles to each other). The plane of a dipole antenna's wires, for example, defines the plane of electric field oscillation, while the magnetic field will be in a plane perpendicular to the wires. Antenna conductors should therefore be *parallel* to each other in order to maximize reception, in order that the electric and magnetic fields emanating from the wires of the transmitting antenna will "link" properly with the wires of the receiving antenna(s). If the goal is optimum communication in any direction (omnidirectionality), dipole and whip antennas should be arranged *vertically* so that all antenna conductors will be parallel to each other regardless of their geographic location.

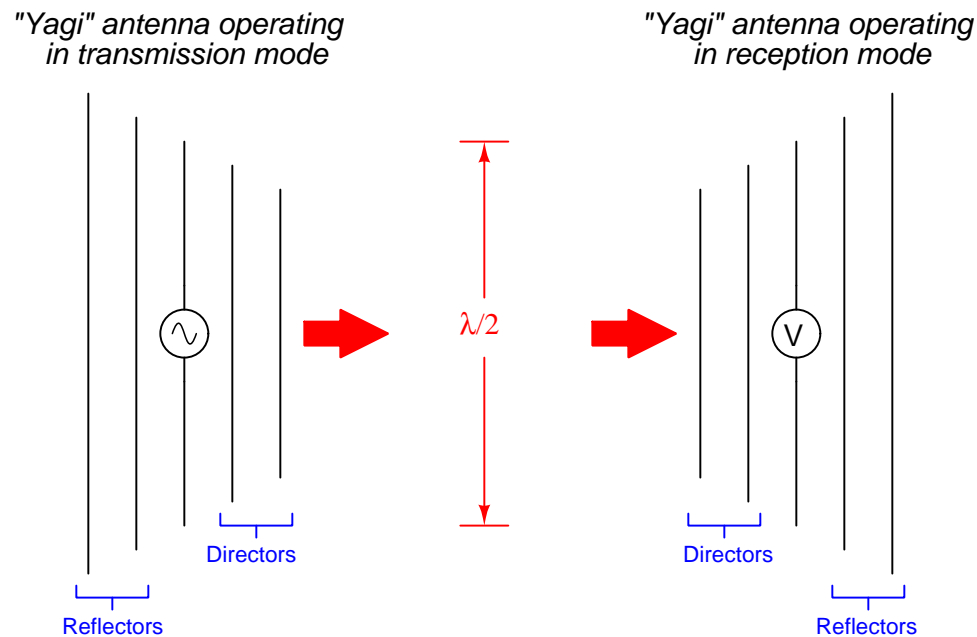
By convention, it is the plane of the *electric field* and not the magnetic field in an electromagnetic wave that defines its polarization. Thus, for a dipole with vertical elements (i.e. antenna wires pointed straight up and down) the emitted radio wave polarization is said to be vertical, and that antenna will be maximally sensitive to received radio waves with that same vertical polarization.

Omnidirectionality may seem like a universally good trait for any antenna: to be able to transmit and receive electromagnetic waves equally well in any direction. However, there are good reasons for wanting directionality in an antenna.

One reason for wanting antenna directionality is improved transmission and reception. As noted before, the AC signal received at an antenna is very small, typically on the order of *microVolts*. Since electromagnetic radiation tends to "spread" as it travels, becoming weaker with distance from the transmitting antenna, long-range radio communication benefits from increased sensitivity. A transmitting antenna that is directional will focus more of its power in one direction than in others, resulting in less "wasted" power radiated in unwanted directions. Likewise, a receiving antenna that is directional does a better job of "collecting" the electromagnetic energy it receives from that direction (as well as ignoring electromagnetic waves coming from other directions). We generally refer to the degree that an antenna focuses or collects electromagnetic waves as the *gain* of that antenna: how much stronger its radiated energy will be in a particular direction or its received signal will be from a wave arriving from a particular direction compared to if it were perfectly omnidirectional.

Another reason for wanting antenna directionality is for greater security: perhaps you have an application where you do *not* wish to broadcast information in all directions, where anyone at all could receive that information. In that case, the best antennas to use would be those that work best in one direction and one direction only, with transmitting and receiving antenna pairs pointed directly at each other.

A simple yet highly effective antenna design for directional transmission and reception is the *Yagi*, named after its inventor Hidetsugu Yagi. A Yagi is based on a half-wave dipole element surrounded by one or more wires longer than $\lambda/2$ to the rear (called “reflectors”) and one or more wires shorter than $\lambda/2$ to the front (called “directors”). The terms “reflector” and “director” are quite apt, describing their interaction with electromagnetic waves from the perspective of the dipole: reflectors *reflect* the waves, while directors *direct* the waves. The result is an antenna array that is much more directional (i.e. will have a greater *gain* than) than a simple dipole:



Yagi antennas in particular excel in providing high gain/directionality. It is quite practical to construct a Yagi antenna having a gain of 20 dB, which means it is *100 times more effective* in transmitting and receiving electromagnetic waves in one direction than a perfectly omnidirectional antenna would be in that same direction.

An example of a Yagi antenna used as part of a municipal wastewater SCADA system is shown in this photograph, the antenna of course being the multi-element array in the upper-right corner:



This particular antenna is part of a radio communication network between “pumping stations” in a city, these stations acting to pump wastewater from gravity-fed collection sumps to the centralized treatment facility.

Another example of a highly directional antenna design is the *parabolic dish*, often used in microwave and satellite communications. This photograph shows two “dish” antennas, one open to the weather (right) and the other covered with a weather-proof diaphragm to keep the antenna elements protected (left):



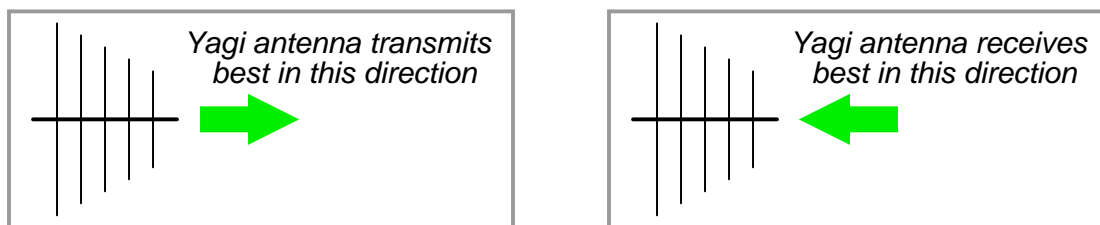
Some dish antennas use mesh or metal-tube reflectors rather than a solid parabolic reflector, as is the case in this next photograph:



The degree of gain or directionality for a dish antenna depends greatly on the size of the reflecting dish compared to the wavelength of the transmitted or received electromagnetic waves. The larger the dish-to-wavelength ratio, the greater the gain and therefore the more directional the antenna will be. It is not uncommon for the gain of a parabolic dish antenna to be 50 dB or more, 50 dB representing a ratio of 100,000 times greater directionality (focus) than an omnidirectional antenna!

3.5 Antenna reciprocity

A basic principle in antenna theory called *reciprocity* states that the efficiency of an antenna as a radiator of electromagnetic waves mirrors its efficiency as a collector of electromagnetic waves. In other words, a good transmitting antenna will be a good receiving antenna, and an antenna having a preferred direction of radiation will likewise be maximally sensitive to electromagnetic waves approaching from that same direction. To use a Yagi as an example:



Related to reciprocity is the concept of equivalent orientation between transmitting and receiving antennas for maximum effectiveness. The electromagnetic waves emitted by a transmitting antenna are *polarized* in a particular orientation, with the electric and magnetic fields perpendicular to each other. The same design of antenna will be maximally receptive to those waves if its elements are similarly oriented. A simple rule to follow is that antenna pairs should always be *parallel* to each other in order to maximize reception, in order that the electric and magnetic fields emanating from the wires of the transmitting antenna will “link” properly with the wires of the receiving antenna(s). If the goal is optimum communication in any direction (omnidirectionality), dipole and whip antennas should be arranged *vertically* so that all antenna conductors will be parallel to each other regardless of their geographic location.

3.6 Antenna characteristics

Different antenna designs vary in several important performance metrics, including:

- Radiation pattern
- Input impedance
- Gain
- Feed signal type (balanced or unbalanced)

Radiation pattern is a geometric expression of how well an antenna sends and receives electromagnetic radiation at different directions. No real antenna is perfectly symmetrical in this regard, and in fact it can be a decided advantage to use an antenna with a “preferred” direction of operation.

We generally denote the radiation pattern for any antenna by sketching a shaded area surrounding the antenna in all directions. One should not think that the outer edges of these shaded areas represents some “hard” boundary beyond which no radiation is emitted (or detected). In reality, the radiation patterns extend out to infinity (assuming otherwise empty space surrounding the antenna). Instead, the size of each shaded area simply represents how effective the antenna is in that direction compared to other directions – just like a “polar” style mathematical plot, where the distance from the graph’s center (origin) point to the periphery of the drawn function represents magnitude at that particular angle or direction. In the case of the vertical whip and dipole antennas, for instance, the radiation patterns show us that these antennas have *zero* effectiveness along the vertical (*Y*) axis centerline. To express this in anthropomorphic terms, these antenna designs are “deaf and mute” in those directions where the radiation pattern is sketched having zero radius from the antenna center.

Input impedance or **Feedpoint impedance** for any antenna is the amount of electrical impedance it presents at the location where a cable or transceiver connects to it. This is important because the antenna must connect to a transmitter or receiver circuit, generally through a cable of some sort, and in order to avoid “reflected” signals and other inefficiencies the antenna’s impedance must match (or be transformed to match) with other components in the system.

The **gain** value for any antenna is a measure of how well it concentrates its radiated output in a particular direction versus other possible directions when transmitting, and also (by the principle of antenna *reciprocity*) how well it receives incoming radiation from a particular direction versus other possible directions. Gain is usually expressed as a decibel figure, calculated by comparing radiated field strength (or received field sensitivity) in the direction of maximum effect versus an antenna that is omnidirectional. Sometimes the benchmark antenna that any given antenna is compared against is a so-called *isotropic* antenna which is a purely theoretical device radiating and receiving equally well in every direction. Other times the benchmark is an ideal dipole antenna which is omnidirectional in one plane only.

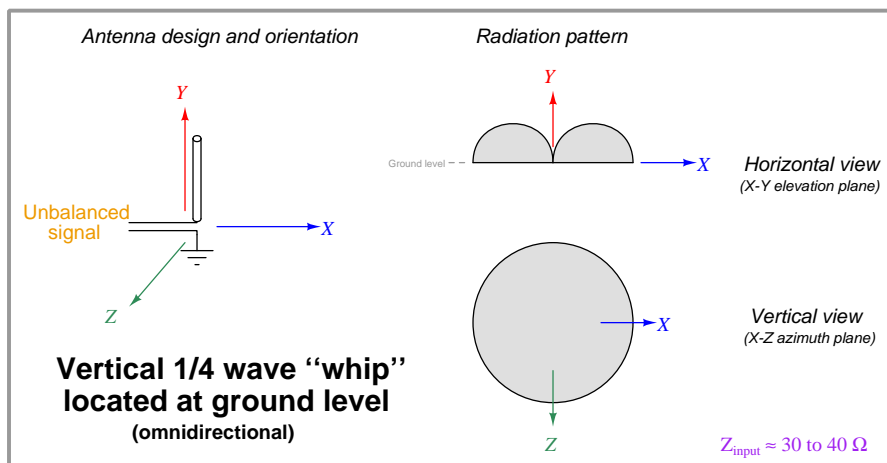
Regarding **feed signal type**, voltage signals may be broadly grouped into two categories: *balanced* and *unbalanced*. All voltages are measured between two points, and if one of those points is Earth ground then the signal is known as “single-ended” or “unbalanced”. When the voltage in question is measured between two points, neither one of which is equipotential to Earth ground,

then the signal is known as “differential” or “balanced”. If an antenna is connected to the wrong type of signal, its radiation resistance, input impedance, and general performance may significantly degrade.

Antenna positioning is an extremely important factor for its performance. The height above the Earth, distance from electrically conductive surfaces, and orientation all affect an antenna’s characteristic measurements and performance. Proximity to any conductive mass in particular (including the Earth) distorts an antenna’s radiation pattern, which in turn affects how well it can transmit and receive in certain directions. If there is any consistent rule to follow when setting up antennas for maximum performance, it is this: *position them as high as possible and as far away from interfering objects as possible!*

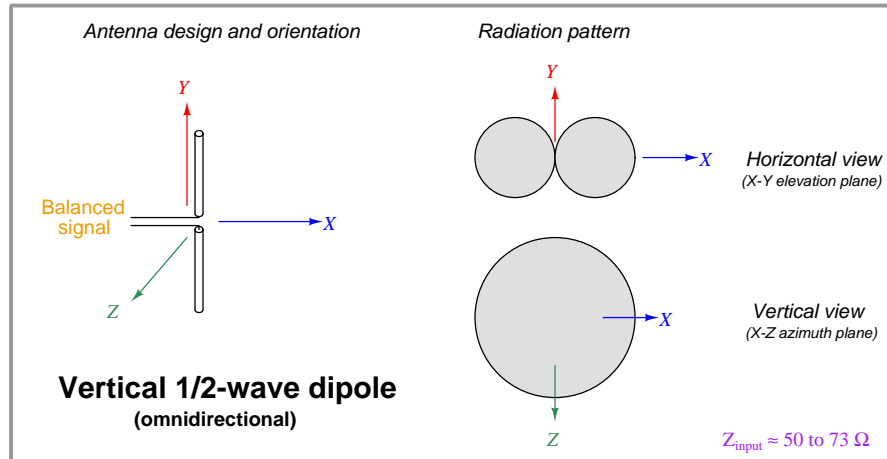
One final note on these antenna types and characteristics is that they are cited for antennas *operating at their fundamental (resonant) frequency*. If an antenna is operated at some frequency substantially different than this, its radiation pattern and input impedance will substantially change!

3.6.1 Vertical monopole antenna



Often called a *whip* antenna, these are quite popular on mobile radio equipment.

3.6.2 Vertical dipole antenna

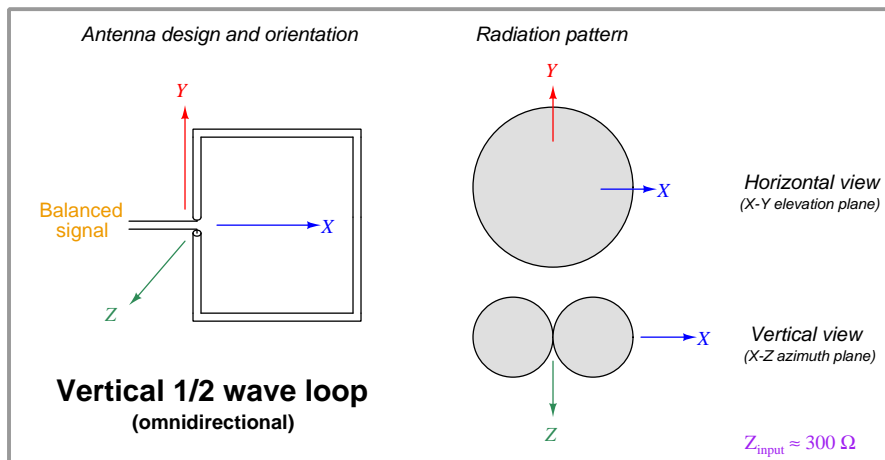


It should be noted that all radiation patterns shown for this and other non-grounded antenna styles assumes the antenna operates within completely free space. The presence of any nearby conductive surfaces will distort the radiation pattern, an extreme example of this being the monopole antenna which is just a dipole with one of the rods buried in Earth ground.

Interestingly, dipole antennas exhibit less gain compared to monopoles, all other factors being equal, because a dipole's radiation pattern extends both below and above the antenna's centerline while the monopole's only extends above the ground plane. This means the monopole radiates into only half the spatial volume that a dipole does, making its radiation pattern stronger in the above-ground region than a vertical dipole's is above centerline.

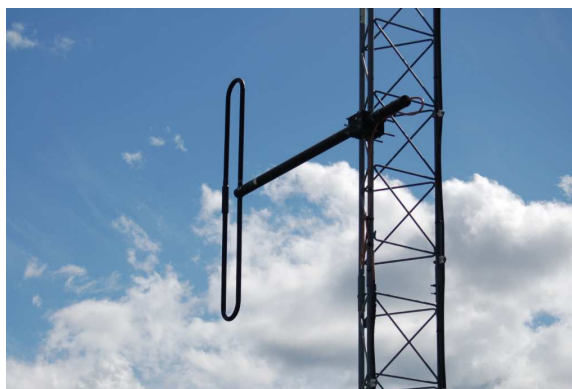
3.6.3 Vertical loop antenna

So far all the antenna designs discussed utilize open-ended conductors for the radiating elements. However, conductive *loops*³ are also capable of radiating and receiving electromagnetic energy:



The radiation pattern for a vertical loop antenna with a perimeter length equal to $\frac{\lambda}{2}$ is orthogonal to that of a half-wave dipole, with maximum gain in the plane of the loop rather than in the plane of Earth ground (i.e. the Z plane). Loop antennas are seldom used because their *loss resistance* to *radiation resistance* ratio is fairly high, which means a greater amount of transmitter power becomes “wasted” in the form of heat than with “open” types of antennas.

One particular variation of the loop design is called the *folded dipole*, where the loop is rectangular in shape rather than square. If the shorter dimension of this loop is a small fraction of one wavelength, the antenna will have the same basic radiation pattern as a normal (open-ended) dipole:

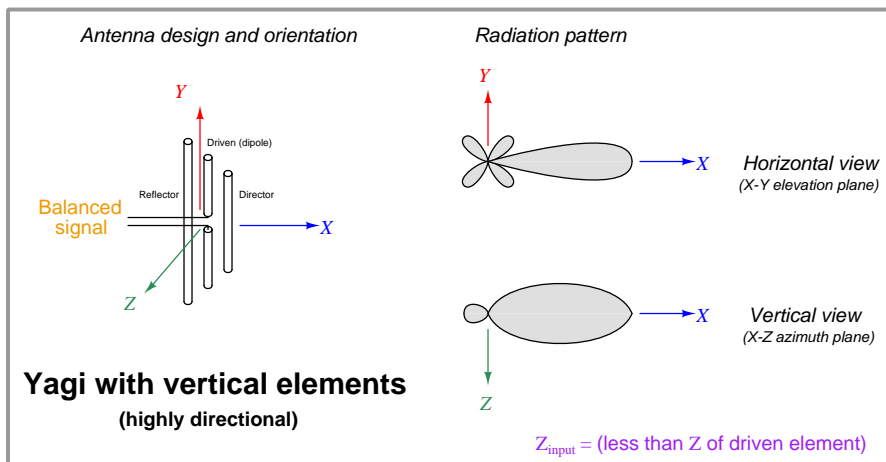


Folded dipole antennas typically exhibit input impedance values approximately four times that of an open-ended dipole, usually around 280 to 300 Ohms.

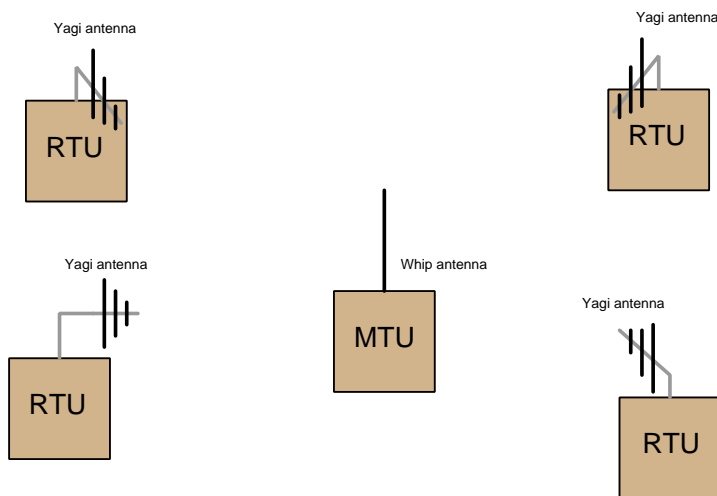
³The actual shape of the loop is less important than one might suspect. Square, circular, and rhombus shapes all function well, with the primary consideration being physical support of the wire loop.

3.6.4 Vertical Yagi antenna

A popular directional antenna design is the *Yagi*, using passive “reflector” and “director” elements to focus the dipole’s sensitivity in one direction:



Yagi antenna pairs may be horizontally or vertically oriented, so long as the transmitting and receiving Yagis are both mounted with the same polarization and face each other. In industrial SCADA radio applications, Yagi antennas are generally oriented with the dipole wires vertical, so that they may be used in conjunction with omnidirectional whip or dipole antennas. An illustration of such use is shown here, with multiple “Remote Terminal Unit” (RTU) transceivers communicating with a central “Master Terminal Unit” (MTU) transceiver:



Here, all the Yagi antennas on the RTUs are vertically oriented, so that they will match the polarization of the MTU’s whip antenna. The Yagi antennas all face in the direction of the MTU for

optimum sensitivity. The MTU – which must broadcast to and receive from all the RTUs – really needs an omnidirectional antenna. The RTUs – which need only communicate with the one MTU and not with each other – work best with highly directional antennas.

If the MTU were equipped with a Yagi antenna instead of a whip, it would only communicate well with one of the RTUs, and perhaps not at all with some of the others. If all RTUs were equipped with whip antennas instead of Yagis, they would not be as receptive to the MTU's broadcasts (lower gain), and each RTU would require greater transmitter power to transmit effectively to the MTU.

Yagi antennas achieve their high directionality (and therefore high gain) by the interaction of the reflector and director elements with the driven element. Each reflector and each director resembles a dipole antenna structure with a short-circuited base, and as such they behave as free-resonating structures. Reflector elements, being longer in length than the driven element, naturally resonate at lower frequencies. Director elements, being shorter, naturally resonate at higher frequencies than the driven element. Both the reflectors and directors receive RF energy from the driven element, and respond to those electromagnetic waves by re-radiating their own waves out of phase with the incident waves since that received energy must be conserved. These phase shifts are analogous to an LC tank circuit being driven by an AC source whose frequency does not match the resonant frequency of that LC network: the network in turn exhibits either an inductive or a capacitive characteristic depending on whether the driving frequency is lower than or greater than resonance.

The genius of the Yagi design is that by careful control of the reflector and director lengths (i.e. their natural resonant frequencies) and the spacings between these elements, it is possible to create a condition where re-radiated waves from each of the director elements happen to fall in-phase with the incident radiation coming from the driven element so as to constructively interfere with that “forward” wave and thereby boost its strength. Meanwhile, the reflector elements' re-radiated waves may be arranged by reflector resonance value(s) and by spacing to destructively interfere with the driven element's incident wave, thereby attenuating the signal strength to the rear. The over-all effect is to focus the antenna's net radiation in one direction as opposed to all other directions.

In addition to achieving directionality, the reflector and director also have the effect of reducing the driven dipole's radiation resistance. A Yagi antenna with a dipole as the driven element may have an input impedance as low as 10 Ohms, much less than the natural radiation resistance of a dipole (approximately 50 to 73 Ohms). Some Yagi antennas therefore employ a *folded dipole* driven element rather than a standard dipole because the folded dipole's impedance is naturally greater than that of a standard dipole, thus compensating for the "loading" effect of the passive elements, and bringing the Yagi antenna's impedance closer to standard coaxial cable impedance and transceiver impedance values (e.g. 50 Ohms). A photograph of a small Yagi antenna with a folded dipole driven element is shown here:



Another compensating strategy for low Yagi antenna impedance is to use something called a *gamma match* between the antenna feedport and a dipole driven element, consisting of an insulated metal rod inside a coaxial metal tube forming a capacitance, the end of this structure attached to one side of the driven dipole. The combination of this series capacitance and the location of the connection point to the driven dipole element forms an impedance-transformation network for RF signals. Two gamma-match networks are visible in the following photograph of a hand-held dual-band (2 meter and 70 centimeter) Yagi antenna intended for amateur radio communication with satellite-based radio systems:



In this antenna the 2-meter (144 MHz to 148 MHz) Yagi array consists of three elements (one director, one driven, and one reflector) while the 70 centimeter (420 MHz to 450 MHz) Yagi array consists of seven elements (five directors, one driven, one reflector). Each gamma-match is clearly visible between its respective coaxial cable and its respective driven element. The two Yagi arrays

are splayed 90 degrees apart along the centerline of the antenna's boom so as to not electrically interfere with each other.

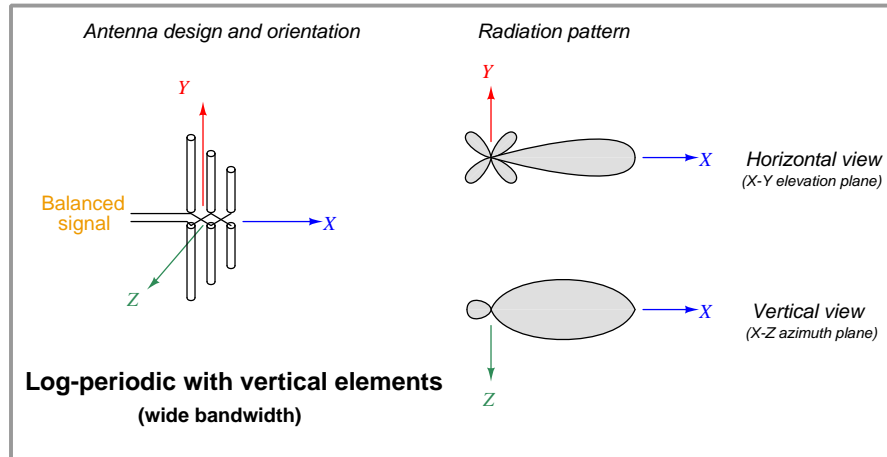
3.6.5 Vertical corner-reflector antenna

Another directional variant of the dipole antenna is the so-called *corner-reflector* antenna, where the driven dipole element is placed in front of a V-shaped reflector array or surface. The following photograph shows a corner-reflector antenna with a folded dipole as the driven element:



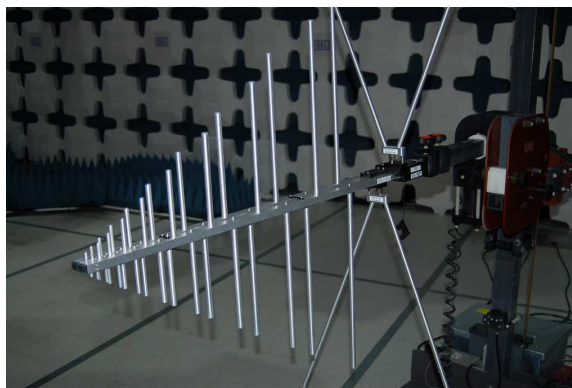
The radiation pattern for a corner reflector antenna is quite similar to that of a Yagi. If a solid metal reflector element is used, the rearward attenuation is quite substantial and typically better (i.e. more directive) than a Yagi though not as good as a parabolic dish.

3.6.6 Vertical log-periodic dipole antenna



An antenna design similar to the Yagi in appearance but different in purpose is the *log-periodic dipole*, with multiple pairs of driven dipole pairs having different lengths and opposing polarities. Unlike the Yagi design which achieves greater antenna gain with more director elements, the log-periodic antenna's purpose is to function well over a broad range of frequencies, that "bandwidth" becoming broader with additional dipole elements.

Log-periodic antennas find application where one antenna must handle the output of a transmitter covering a range of different frequencies, and/or where a receiver unit must function well over a broad range of frequencies using the same antenna. An example of the latter is this log-periodic antenna used inside of an RF test chamber to monitor emitted radio signals from various types of electronic devices. The range of possible emitted signals from an arbitrary selection of electronic devices is quite wide, which is why the log-periodic antenna is a good choice in this application:



If you closely examine this photograph, you will note that the aluminum antenna elements attach alternately to two aluminum "rails" which are the balanced signal conductors, so that each pair of matched-length elements has a polarity opposite its neighboring pairs.

3.6.7 Circularly-polarized antennas

In addition to planar-polarization (typically in either the vertical or horizontal planes) it is also possible to generate electromagnetic waves that propagate with *circular* polarization. This means the orientation of the wave’s electric field actually *rotates* as it travels through space. Circular polarization is useful for applications where the receiving and transmitting antenna orientations are not necessarily fixed relative to each other, such as in the case of satellite or space probe antennas where the craft is likely to rotate or even tumble over time.

Creating a circularly-polarized radio wave is not as difficult as it might seem. The basic concept is not unlike that of the *rotating magnetic field* concept upon which nearly all AC electric motors are designed. One way to create this effect is to use a multi-element antenna where the different driven elements are energized out-of-phase with each other. The following photograph shows a circularly-polarized Yagi antenna⁴ intended for communication with orbiting satellites:



Note the driven elements in this antenna (the rods with black-colored tips), positioned about the boom in an “X” configuration. The incoming feedline for this antenna splits into two coaxial cables, one of which is exactly one-quarter of a wavelength longer⁵ than the other at the antenna’s resonant frequency. This means one dipole’s signal will be phase-shifted by 90 degrees from the other. This is analogous to a two-phase AC electric motor having stator windings splayed 90 degrees from each other and energized by currents that are also 90 degrees phase-shifted from each other.

Other methods also exist for generating circularly polarized radio waves. Some “X” form Yagi antennas stagger the two sets of driven dipoles, reflectors, and directors such that each successive dipole pair is one-quarter of a wavelength further along the antenna boom than the pair before. This way, the two sets of driven elements may be driven in-phase with each other, but their physical locations along the antenna boom provide the “spin” to circularly polarize the traveling wave. Another form of circularly polarized antenna uses a helical-shaped wire element rather than straight dipole elements.

⁴This design is sometimes called a *turnstile antenna* due to its appearance.

⁵The longer cable must be one-quarter wavelength *electrically* longer and not one-quarter wavelength *physically* longer, owing to the velocity factor of the cable slowing down the signal’s speed to below that of the speed of light in air or vacuum.

Though circular polarization eliminates the vertical-versus-horizontal alignment problem of standard (“planar”) antennas, it does introduce a new polarization difference which must be accounted for: *right-hand* versus *left-hand* circular polarization, and the direction of circular polarization must match between the transmitting and receiving antennas in order to ensure good signal integrity. Fortunately, this is easy to accomplish through antenna design, and so long as both antennas have the same circular polarization all will be well.

For example, right-hand circular polarization (RHCP) – also known as clockwise (CW) polarization – works like the threads on a common screw or bolt. If you imagine a standard bolt with “right-hand threads” pointed with its long axis toward the sky, in the same direction as the boom of a directional antenna pointed toward a satellite, as the radio wave travels from the transmitter up the antenna boom and into empty space the electric and magnetic fields will rotate in a clockwise direction as viewed from the ground looking up toward the satellite. This is the same direction that a nut would have to turn on the bolt to travel up toward the sky as well. If the satellite also has a RHCP antenna, the received signal’s rotation will mesh with the satellite’s antenna like a nut being threaded onto (“received by”) another standard bolt with “right-hand threads” on the satellite pointed toward Earth. However, if the Earth station used a RHCP and the satellite a LHCP antenna, the received radio wave will no more be received by that antenna any more than a right-hand-threaded nut would thread itself onto a left-hand-threaded bolt if launched toward it from the ground.

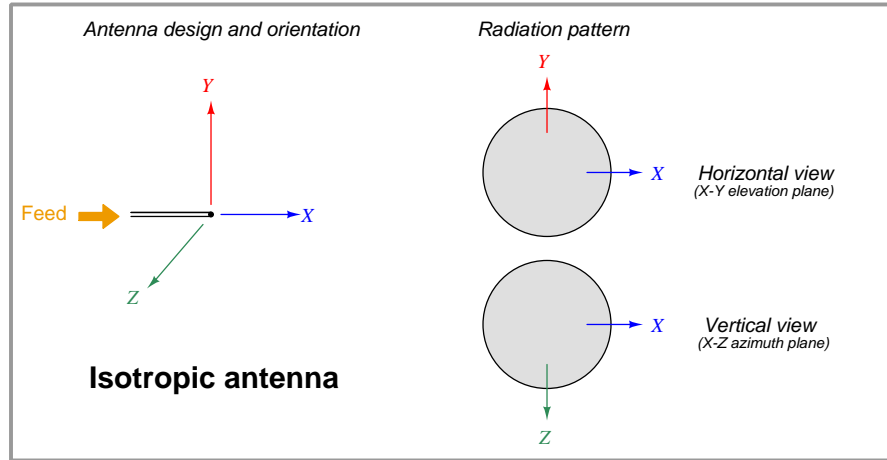
It is worth noting that circular-polarized antennas and planar-polarized antennas are able to communicate with each other, just not as strongly as if the transmitting and receiving antennas match each other’s polarization types. In such circumstances the rotational angle of the planar antenna becomes irrelevant so long as it still points toward the satellite!

3.7 Antenna gain calculations

A common way to express the maximal effectiveness of any antenna design is as a ratio compared to some idealized form of antenna with a more uniform radiation pattern. As with most ratio measurements in radio technology, the standard unit for antenna gain is the *decibel* (dB), related to a ratio of powers as follows:

$$\text{Gain in dB} = 10 \log \left(\frac{P}{P_{ref}} \right)$$

Since antenna gain is not a real amplification of power – this being impossible according to the Law of Energy Conservation – greater antenna gain is achieved only by greater *focus* of the radiation pattern in a particular direction. Therefore, antenna gain is really just another way of stating how well it focuses in one particular direction. The most common reference standard used to calculate antenna gain is a purely theoretical device called an *isotropic antenna*. This is an ideally omnidirectional antenna having a perfectly spherical radiation pattern, radiating energy equally in all directions:



If a directional antenna such as a Yagi radiates (and/or receives) 20 times as much power in its most sensitive direction as an isotropic antenna, it is said to have a power gain of 13.01 dBi (13.01 *decibels* more than an *isotropic*). An alternative “reference” for comparison is a half-wave dipole antenna. Decibel comparisons against a dipole are abbreviated dBd. The assumption of an isotropic antenna as the reference is so common in radio engineering, though, you often see antenna gains expressed simply in units of dB. The assumption of isotropic reference (dBi) for antenna gain expressions is analogous to the assumption of “RMS” measurements in AC circuits rather than “peak” or “peak-to-peak”: if you see an AC voltage expressed without any qualifier (e.g. “117 Volts AC”), it is generally assumed to be an RMS measurement.

The theoretical gain of a dipole antenna is 2.15 dBi, which means its radiated signal strength in the direction of maximum focus will be 2.15 dB or 1.64 times greater than that of an ideal isotropic antenna. Monopole (“whip”) antennas optimally exhibit a gain of 6 dBi (6 dB being approximately equal to a 4-fold magnification compared to an isotropic antenna), while Yagis typically achieve 15

dBi or more depending on the number of director elements. The reason why monopole antennas exhibit greater gain than dipoles is because their radiation pattern is concentrated only in the space *above* ground level whereas dipoles radiate both above and below their horizontal centerlines, but this comparison holds true only for ideal conditions where the monopole operates above a perfectly conductive ground plane and the dipole operates in perfectly free space away from all conductive surfaces. Real whip antenna performance varies significantly with installation. Parabolic “dish” antenna designs such as those used in microwave communication systems may achieve gains of 50 dBi ($\approx 100,000$ -fold magnification) or more.

The flip-side of high-gain antennas, loudspeakers, and microphones is how poorly they perform in directions other than their preferred direction. Any transmitting or receiving structure exhibiting a “gain” due to its focused radiation pattern must likewise exhibit a “loss” in performance when tested in directions other than its direction of focus. Referring back to directional radio antennas again, a Yagi with an advertised gain of 15 dBi (in its “preferred” direction) will exhibit a strong negative gain in the rearward direction where its ability to radiate and receive is almost non-existent.

Measurements of *front-to-back ratio* are often made for directional antennas such as the Yagi to quantify how much better they are at sending and receiving in their forward direction compared to their rearward direction. A practical contrast of two antenna types having very different front-to-back ratios are the Yagi versus the log-periodic. These two antenna types look remarkably similar at first glance but are actually constructed very differently from one another, the Yagi consisting of a simple dipole (or sometimes a folded dipole) driven element with one reflector element behind and one or more director elements ahead of the driven element, and the log-periodic consisting of all driven elements (or differing length) with alternating polarities. Yagis are optimized for maximum directionality (gain), while log-periodics are optimized for good performance over a broad range of frequencies (i.e. performance over a large bandwidth). While still exhibiting higher gain than a dipole antenna, a log-periodic does worse than any Yagi having the same number of elements, and we see this lesser directionality mostly in the poorer *front-to-back* ratio of the log-periodic compared to the Yagi. In other words, with all other factors being equal, a Yagi projects more of its radiated energy in one direction (front), whereas the log-periodic splits more between projecting energy out the front and out the rear.

Parabolic dish antennas exhibit some of the greatest gain values of any antenna design, meaning they are the most directional. For such antennas the gain may be estimated by the following formula:

$$\text{Gain} = 10 \log k \left(\frac{\pi D}{\lambda} \right)^2$$

Where,

Gain = antenna gain in decibels (greater than an isotropic antenna)

k = Efficiency factor⁶ of the antenna, typically 0.5 to 0.6

D = Diameter of the parabolic dish, in the same unit of measurement as wavelength

λ = Signal wavelength

If even more signal gain is necessary than what may be achieved by narrower radiation focus, an actual electronic amplifier may be added to an antenna assembly to boost the RF power sent to

⁶This factor depends on several variables, including the illumination area of the dish, the reflectivity of the dish’s surface to electromagnetic waves, etc.

or received from the antenna. This is common for satellite dish-type receiving antennas, where a “low-noise amplifier” or *LNA* is often located right at the focal point of the parabolic dish. Satellite communication requires very high transmitter and receiver gains, due to inevitable signal weakening over the extremely long distances between a ground-based antenna and a satellite antenna in geosynchronous orbit around the Earth.

3.8 Antenna standing wave ratio

Another factor affecting the performance of any antenna connected to a radio transmitter is the *standing wave ratio*. Recall from your study of transmission lines that all cables exhibit characteristic impedance, and if that surge impedance of a cable does not match the impedance of whatever device(s) connect to its end the result will be energy reflected back toward the source after the traveling wave front reaches the transition point between the cable and the non-matching impedance. The constructive and destructive interference of the advancing waves with the reflected waves creates areas along the cable with maximum voltage or maximum current (anti-nodes) as well as areas along the cable with minimum voltage or minimum current (nodes), which appears as though there is a stationary or *standing* wave existing on that cable.

Antennas are supposed to act as loads to any signal energy arriving to them from a cable, but if the characteristic impedance of that feedline cable does not match the impedance of the antenna, some of the energy will reflect back toward the source rather than go into the antenna to be radiated into space as electromagnetic waves. Excessive reflections resulting from cable/antenna impedance mis-match in a radio transmitter system results in less energy being radiated by the antenna and more energy dissipated along the cable and also within the transmitting circuitry, sometimes enough energy even to damage the transmitter's final driving devices (transistors or vacuum tubes)!

An expression of a standing wave's maximum versus minimum amplitudes is called the *standing wave ratio* or *SWR*. Practically speaking, these measurements are usually expressed in terms of voltage and are most properly called VSWR. An ideal VSWR measurement is 1:1 which means the peak voltage measured at any point along the cable is equal to the peak voltage measured anywhere else on that same cable – i.e. there are no standing waves, no nodes or anti-nodes.

Standing wave ratio, or *SWR*, may be calculated from measurements of forward and reflected power taken using a directional RF wattmeter connected between the transmitter and antenna. A more convenient instrument to use, however, is a special-purpose *SWR meter*. An example of an SWR meter is shown in the photograph below, taken of the faceplate of a Citizen's Band (CB) radio transceiver:



In order to improve transmitting efficiency and also to minimize damage to the radio transmitter from reflected power, a common technique to adjust VSWR to a value closer to unity (1) is to alter the physical length of the antenna. To this end, antenna elements may be built to be telescoping. “Trimming” the length of a transmitting antenna, therefore, consists of monitoring the SWR meter while making test transmissions at slightly different frequencies⁷.

⁷On a Citizen's Band transmitter, the “channel” number is proportional to frequency, with channel 1 being 26.965 MHz and channel 40 being 27.405 MHz. If SWR gradually grows worse (i.e. increases) as frequency is increased across the entire CB band it means the antenna is too long and needs to be shortened. Ideally the antenna's length should be adjusted until the best SWR measurement is obtained at a frequency in the middle of the band (or, if you don't intend to use the entire band then at whatever channel within the band you plan to use most).

Chapter 4

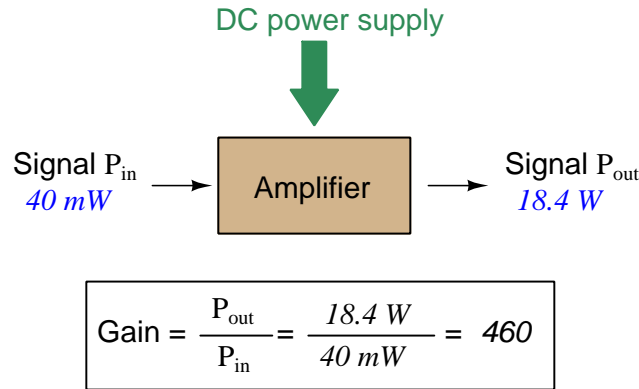
Derivations and Technical References

This chapter is where you will find mathematical derivations too detailed to include in the tutorial, and/or tables and other technical reference material.

4.1 Decibels

One of the mathematical tools popularly used to gauge increases and decreases of electrical power is the *common logarithm*, expressed as a measurement unit called the *decibel*. The basic idea of decibels is to express a ratio of two electrical power quantities in logarithmic terms. Every time you see the unit of “decibel” you can think: *this is an expression of how much greater (or how much smaller) one power is to another*. The only question is which two powers are being compared.

Electronic amplifiers are a type of electrical system where comparisons of power are useful. Students of electronics learn to compare the output power of an amplifier against the input power as a unitless ratio, called a *gain*. Take for example an electronic amplifier with a signal input of 40 milliWatts and a signal output of 18.4 Watts:



An alternative way to express the gain of this amplifier is to do so using the unit of the *Bel*, defined as the common logarithm of the gain ratio:

$$\log \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right) = \log \left(\frac{18.4 \text{ W}}{40 \text{ mW}} \right) = 2.66276 \text{ B}$$

When you see an amplifier gain expressed in the unit of “Bel”, it’s really just a way of saying “The output signal coming from this amplifier is x powers of ten greater than the input signal.” An amplifier exhibiting a gain of 1 Bel outputs 10 times as much power as the input signal. An amplifier with a gain of 2 Bels boosts the input signal by a factor of 100. The amplifier shown above, with a gain of 2.66276 Bels, boosts the input signal 460-fold.

At some point in technological history it was decided that the “Bel” (B) was too large and cumbersome, and so it became common to express powers in fractions of a Bel instead: the *deci*Bel (1 dB = $\frac{1}{10}$ of a Bel). Therefore, this is the form of formula you will commonly see for expressing electrical signal power gains or losses:

$$\text{dB} = 10 \log \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right)$$

The gain of our hypothetical electronic amplifier, therefore, would be more commonly expressed as 26.6276 dB rather than 2.66276 B, although either expression is technically valid¹.

¹It is interesting to note that although the “Bel” is a metric unit, it is seldom if ever used without the metric prefix

An operation students often struggle with is converting a decibel figure back into a ratio, since the concept of logarithms seems to be universally perplexing. Here I will demonstrate how to algebraically manipulate the decibel formula to solve for the power ratio given a dB figure.

First, we will begin with the decibel formula as given, solving for a value in decibels given a power ratio:

$$\text{dB} = 10 \log(\text{Ratio})$$

If we wish to solve for the ratio, we must “undo” all the mathematical operations surrounding that variable. One way to determine how to do this is to reverse the order of operations we would follow if we knew the ratio and were solving for the dB value. After calculating the ratio, we would then take the logarithm of that value, and then multiply that logarithm by 10: start with the ratio, then take the logarithm, then multiply last. To un-do these operations and solve for the ratio, we must un-do each of these operations in reverse order. First, we must un-do the multiplication (by dividing by 10):

$$\frac{\text{dB}}{10} = \frac{10 \log(\text{Ratio})}{10}$$

$$\frac{\text{dB}}{10} = \log(\text{Ratio})$$

Next, we must un-do the logarithm function by applying its mathematical inverse to both sides of the formula – making each expression a power of 10:

$$10^{\frac{\text{dB}}{10}} = 10^{\log(\text{Ratio})}$$

$$10^{\frac{\text{dB}}{10}} = \text{Ratio}$$

To test our algebra, we can take the previous decibel value for our hypothetical amplifier and see if this new formula yields the original gain ratio:

$$\text{Ratio} = 10^{\frac{26.6276 \text{ dB}}{10}}$$

$$\text{Ratio} = 10^{2.66276} \text{ B}$$

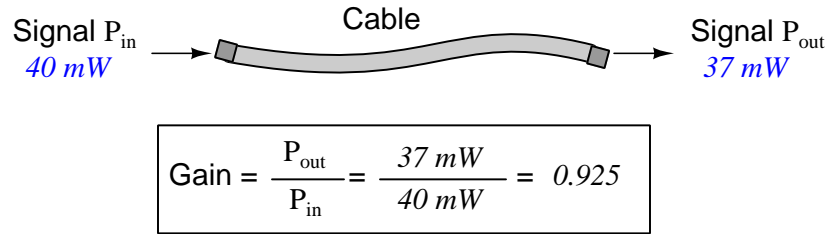
$$\text{Ratio} = 460$$

Sure enough, we arrive at the correct gain ratio of 460, starting with the decibel gain figure of 26.6276 dB.

“deci” ($\frac{1}{10}$). One could express powers in microbels, megabels, or any other metric prefix desired, but it is never done in industry: only the decibel is used.

We may also use decibels to express power *losses* in addition to power *gains*. There are many practical applications of this in signaling systems, both electronic and optical. One such application is *filtering*, where a “filter” circuit screens out certain components of the signal while letting others pass through (e.g. the bass or treble control for an audio system). Another application is *attenuation*, where the entirety of a signal is reduced in magnitude (e.g. the volume control for an audio system).

We will explore yet another application of signal power reduction as a case study for decibels: *cable loss*. Cables designed to convey signals over long distances are not perfect conduits of energy, as some of the signal’s energy is inevitably lost along the way. This is true for different types of signals, electrical and optical being two popular examples. In the following illustration we see a signal cable losing power along its length², such that the power out is less than the power in:



$$10 \log \left(\frac{P_{out}}{P_{in}} \right) = 10 \log \left(\frac{37\text{ mW}}{40\text{ mW}} \right) = -0.3386\text{ dB}$$

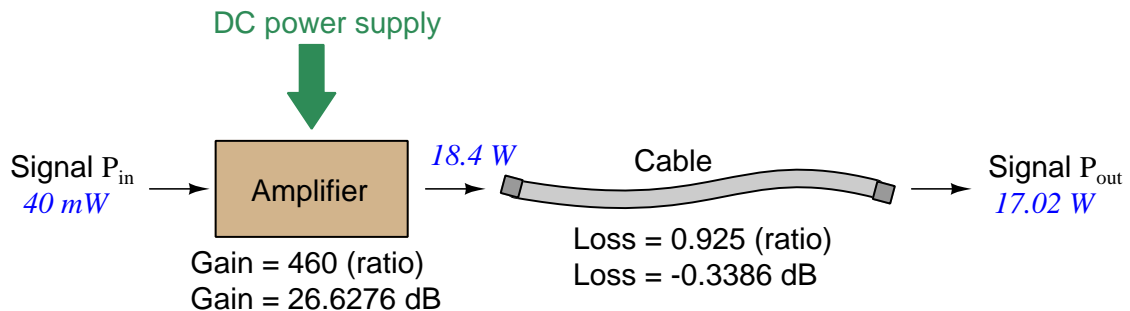
Contrasting this result against the previous result (with the amplifier) we see a very important property of decibel figures: any power *gain* is expressed as a *positive* decibel value, while any power *loss* is expressed as a *negative* decibel value. Any component outputting the exact same power as it takes in will exhibit a “gain” value of 0 dB (equivalent to a gain *ratio* of 1).

Remember that Bels and decibels are nothing more than logarithmic expressions of “greater than” and “less than”. Positive values represent powers that are *greater* while negative values represent powers that are *lesser*. Zero Bel or decibel values represent *no change* (neither gain nor loss) in power.

A couple of simple decibel values are useful to remember for approximations, where you need to quickly estimate decibel values from power ratios (or vice-versa). Each addition or subtraction of 10 dB exactly represents a 10-fold multiplication or division of power ratio: e.g. +20 dB represents a power ratio gain of $10 \times 10 = 100$, whereas −30 dB represents a power ratio reduction of $\frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{1000}$. Each addition or subtraction of 3 dB approximately represents a 2-fold multiplication or division of power ratio: e.g. +6 dB is approximately equal to a power ratio gain of $2 \times 2 = 4$, whereas −12 dB is approximately equal to a power ratio reduction of $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$. We may combine ± 10 dB and ± 3 dB increments to come up with ratios that are products of 10 and 2: e.g. +26 dB is approximately equal to a power ratio gain of $10 \times 10 \times 2 \times 2 = 400$.

²For high-frequency signals such as those used in radio communications, the dominant mode of energy dissipation is *dielectric heating*, where the AC electric field between the cable conductors excites the molecules of the conductor insulation. This energy loss manifests as heat, which explains why there is less signal energy present at the load end of the cable than is input at the source end of the cable. For DC and low-frequency AC circuits the dominant mode of energy dissipation is cable conductor resistance, which is typically very small.

Observe what happens if we combine a “gain” component with a “loss” component and calculate the overall power out versus power in:



The overall gain of this amplifier and cable system expressed as a ratio is equal to the *product* of the individual component gain/loss ratios. That is, the gain ratio of the amplifier *multiplied* by the loss ratio of the cable yields the overall power ratio for the system:

$$\text{Overall gain} = \frac{17.02 \text{ W}}{40 \text{ mW}} = (460)(0.925) = 425.5$$

The overall gain may be alternatively expressed as a decibel figure, in which case it is equal to the *sum* of the individual component decibel values. That is, the decibel gain of the amplifier *added* to the decibel loss of the cable yields the overall decibel figure for the system:

$$\text{Overall gain} = 10 \log \left(\frac{17.02 \text{ W}}{40 \text{ mW}} \right) = 26.6276 \text{ dB} + (-0.3386 \text{ dB}) = 26.2890 \text{ dB}$$

It is often useful to be able to estimate decibel values from power ratios and vice-versa. If we take the gain ratio of this amplifier and cable system (425.5) and round it down to 400, we may easily express this gain ratio as an expanded product of 10 and 2:

$$425.5 \approx 400 = (10) \times (10) \times (2) \times (2)$$

Knowing that every 10-fold multiplication of power ratio is an addition of +10 dB, and that every 2-fold multiplication of power is an addition of +3 dB, we may express the expanded product as a sum of decibel values:

$$(10) \times (10) \times (2) \times (2) = (10 \text{ dB}) + (10 \text{ dB}) + (3 \text{ dB}) + (3 \text{ dB}) = 26 \text{ dB}$$

Therefore, our power ratio of 425.5 is approximately equal to +26 decibels.

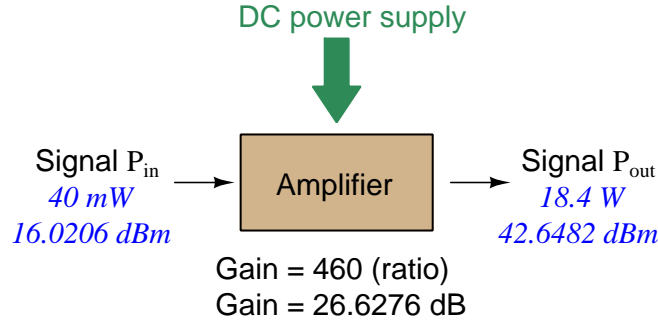
Decibels always represent comparisons of power, but that comparison need not always be P_{out}/P_{in} for a system component. We may also use decibels to express an amount of power compared to some standard reference. If, for example, we wished to express the input power to our hypothetical amplifier (40 milliWatts) using decibels, we could do so by comparing 40 mW against a standard “reference” power of exactly 1 milliWatt. The resulting decibel figure would be written as “dBm” in honor of the 1 milliWatt reference:

$$P_{in} = 10 \log \left(\frac{40 \text{ mW}}{1 \text{ mW}} \right) = 16.0206 \text{ dBm}$$

The unit of “dBm” literally means the amount of dB “greater than” 1 milliWatt. In this case, our input signal of 40 milliWatts is 16.0206 dB greater than a standard reference power of exactly 1 milliWatt. The output power of that amplifier (18.4 Watts) may be expressed in dBm as well:

$$P_{out} = 10 \log \left(\frac{18.4 \text{ W}}{1 \text{ mW}} \right) = 42.6482 \text{ dBm}$$

A signal power of 18.4 Watts is 42.6482 dB greater than a standard reference power of exactly 1 milliWatt, and so it has a decibel value of 42.6482 dBm.



Notice how the output and input powers expressed in dBm relate to the power gain of the amplifier. Taking the input power and simply *adding* the amplifier’s gain factor yields the amplifier’s output power in dBm:

$$P_{in}(\text{dB}) + P_{gain}(\text{dB}) = P_{out}(\text{dB})$$

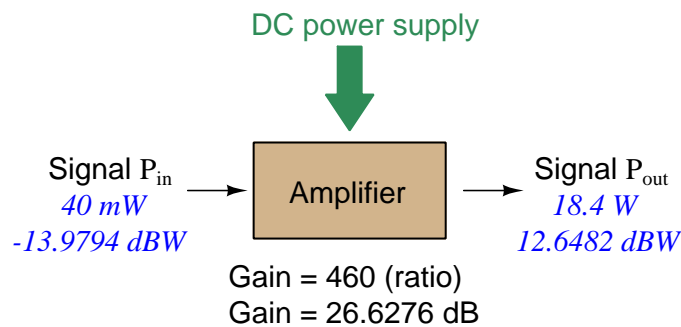
$$16.0206 \text{ dBm} + 26.6276 \text{ dB} = 42.6482 \text{ dBm}$$

An electronic signal that begins 16.0206 dB greater than 1 milliWatt, when boosted by an amplifier gain of 26.6276 dB, will become 42.6482 dB greater than the original reference power of 1 milliWatt.

We may alternatively express all powers in this hypothetical amplifier in reference to a 1-Watt standard power, with the resulting power expressed in units of “dBW” (decibels greater than 1 Watt):

$$P_{in} = 10 \log \left(\frac{40 \text{ mW}}{1 \text{ W}} \right) = -13.9794 \text{ dBW}$$

$$P_{out} = 10 \log \left(\frac{18.4 \text{ W}}{1 \text{ W}} \right) = 12.6482 \text{ dBW}$$



Note how the input power of 40 milliWatts equates to a negative dBW figure because 40 milliWatts is *less* than the 1 Watt reference, and how the output power of 18.4 Watts equates to a positive dBW figure because 18.4 Watts is *more* than the 1 Watt reference. A positive dB figure means “more than” while a negative dB figure means “less than.”

Note also how the output and input powers expressed in dBW still relate to the power gain of the amplifier by simple addition, just as they did when previously expressed in units of dBm. Taking the input power in units of dBW and simply *adding* the amplifier’s gain factor yields the amplifier’s output power in dBW:

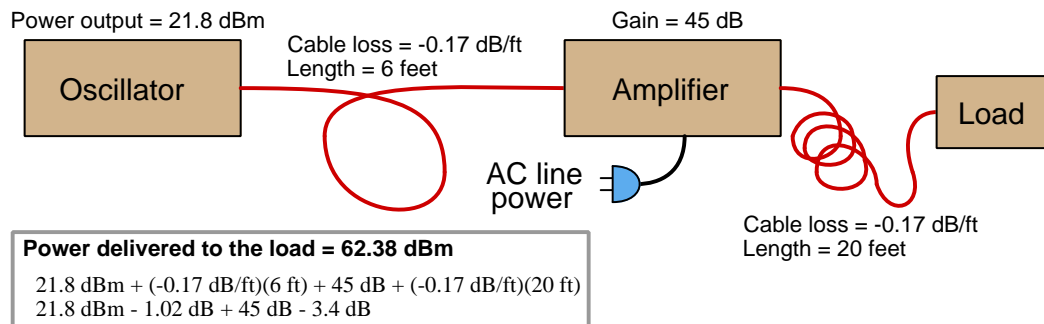
$$P_{in}(\text{dB}) + P_{gain}(\text{dB}) = P_{out}(\text{dB})$$

$$-13.9794 \text{ dBW} + 26.6276 \text{ dB} = 12.6482 \text{ dBW}$$

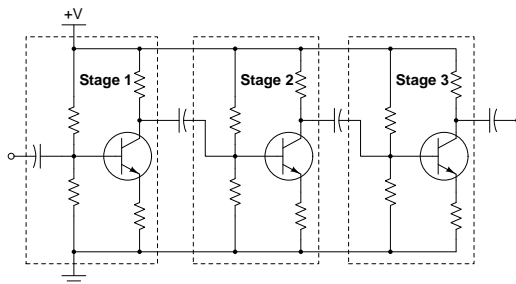
An electronic signal that begins 13.9794 dB less than 1 Watt, when boosted by an amplifier gain of 26.6276 dB, will become 12.6482 dB greater than the original reference power of 1 Watt.

This is one of the major benefits of using decibels to express powers: we may very easily calculate power gains and losses by summing a string of dB figures, each dB figure representing the power gain or power loss of a different system component. Normally, any compounding of *ratios* involves multiplication and/or division of those ratios, but with decibels we may simply add and subtract. One of the interesting mathematical properties of logarithms is that they “transform”³ one type of problem into a simpler type: in this case, a problem of multiplying ratios into a (simpler) problem of adding decibel figures.

For example, we may express the power dissipated along a cable in terms of decibels per foot; the longer the cable, of course, the more power will be lost this way, all other factors being equal. For example, a radio-frequency signal cable having a loss figure of -0.15 decibels per foot at a signal frequency of 2.4 GHz will suffer -15 dB over 100 feet, and -150 dB over 1000 feet. To illustrate how decibels may be used to calculate power delivered to a load in such a system, accounting for various gains and losses along the way using decibel figures:



A similar application of decibels is found in multi-stage amplifier circuits, where one stage amplifies a signal to be fed into a successive stage to be amplified more. The power gains of these stages, each expressed as a ratio, *multiply* to make the over-all amplifier's power gain (ratio). The power gains of those same stages, each expressed as a decibel figure, *add* to make the over-all amplifier's power gain (dB):



³In fact, logarithms are one of the simplest examples of a *transform function*, converting one type of mathematical problem into another type. Other examples of mathematical transform functions used in engineering include the *Fourier transform* (converting a time-domain function into a frequency-domain function) and the *Laplace transform* (converting a differential equation into an algebraic equation).

Another common application of decibels is to express ratios of voltage and/or current rather than power. However, since the unit of the Bel has been defined as an expression of a *power* ratio, we cannot simply substitute V or I for P in any of the formulae we've seen so far.

Suppose an amplifier has a voltage gain of 2 (i.e. V_{out} is twice as large as V_{in}), and we would like to express this gain in decibels. Since decibels are intended to express power gain and not voltage gain, we must figure out how much power gain is equivalent to a voltage gain of two. Obviously, voltage and power are fundamentally different quantities, but if we imagine ourselves connecting a fixed load resistance to the input signal, and then to the output signal, we will realize that load's power dissipation will be more than double when energized by a voltage twice as large. Joule's Law is helpful to determine the exact ratio of power dissipation:

$$P = \frac{V^2}{R}$$

Doubling voltage for any given load resistance results in power *quadrupling* because power is proportional to the square of the voltage applied to a fixed resistance. Using this as the basis for applying decibels to a voltage ratio. Knowing that Joule's Law also declares power is proportional to the square of the current applied to a fixed resistance ($P = I^2 R$) means this same mathematical relationship will apply to current gains and reductions as well as voltage gains and reductions:

$$\text{dB} = 10 \log \left(\frac{P_{out}}{P_{in}} \right) = 10 \log \left(\frac{V_{out}}{V_{in}} \right)^2 = 10 \log \left(\frac{I_{out}}{I_{in}} \right)^2$$

An algebraic identity of logarithms is that the logarithm of any quantity raised to a power is equal to that power multiplied by the logarithm of the quantity. Expressed in general terms:

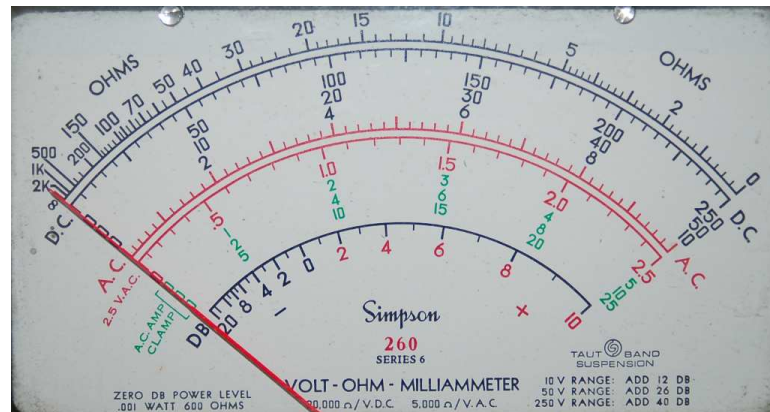
$$\log x^y = y \log x$$

Therefore, we may simplify the decibel formula for voltage gain by removing the "2" power and making it a multiplier:

$$\begin{aligned} 10 \log \left(\frac{V_{out}}{V_{in}} \right)^2 &= (2)(10) \log \left(\frac{V_{out}}{V_{in}} \right) = 20 \log \left(\frac{V_{out}}{V_{in}} \right) \\ 10 \log \left(\frac{I_{out}}{I_{in}} \right)^2 &= (2)(10) \log \left(\frac{I_{out}}{I_{in}} \right) = 20 \log \left(\frac{I_{out}}{I_{in}} \right) \end{aligned}$$

Thus, we may use decibels to express voltage or current ratios if we simply substitute 20 instead of 10 as the multiplier.

We can see the practicality of using decibels to represent something other than electrical *power* by examining this analog meter face, belonging to a Simpson model 260 VOM (Volt-Ohm-Milliammeter). Note the bottom scale on this meter's face, calibrated in decibels (DB):



Pay attention to the note on decibels written in the lower-left corner of the meter face, where 0 dB is defined as 0.001 Watt dissipated by 600 Ohms. The fact that 0 dB is defined as 1 milliWatt means it should (properly) be labeled dBm rather than dB⁴. A load resistance value is necessary as part of this definition for dB because this meter cannot measure power directly but must infer signal power from measurements of AC *voltage*. Without a specific load resistance, there is no clear relation between voltage and power. 600 Ohms is an old telecommunications standard for audio-frequency AC signals, and continues to be used today for voltage-based decibel measurements of audio-frequency AC signals.

The meter as shown is connected to nothing at all, and so registers 0 Volts AC. This, of course, corresponds to zero power, and it has no corresponding decibel value because the logarithm of zero is mathematically *undefined*⁵. Practically, it means $-\infty$ dB, which is why the needle at the 0 Volt position “falls off” the left-hand end of the dB scale.

Close inspection of the dB scale on this meter face reveals another interesting property of decibels, and that is the *nonlinear* nature of the dB scale. This contrasts starkly against all the voltage and current scales on this meter face which are linear. This nonlinearity is a fundamental property of decibels because it is based on the *logarithm* function which is nonlinear.

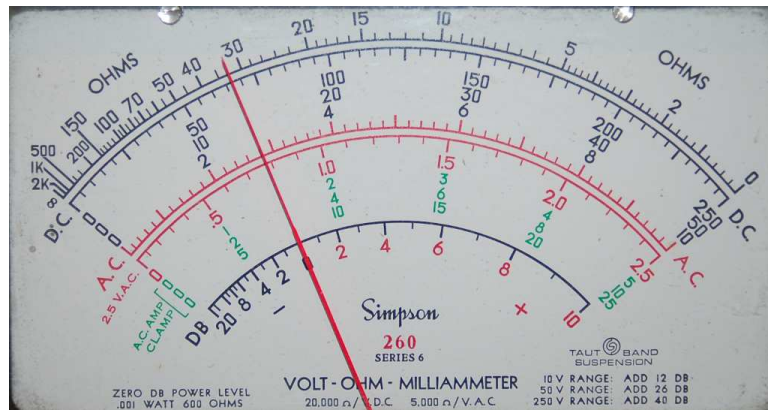
⁴Such mis-labeling is not that uncommon in the profession, the expectation being that the technician or engineer working with the instrument ought to be familiar enough with the concept of decibels to know when dB really means dBm, or dBW, etc.

⁵Your electronic calculator will complain if you attempt to take the logarithm of zero!

Now, we will explore what is necessary to make this meter register 0 dBm (i.e. 1 milliWatt) with an applied AC voltage. 1 milliWatt of power dissipated by 600 Ohms is equivalent to:

$$V = \sqrt{PR} = \sqrt{(0.001)(600)} = 0.7746 \text{ Volts}$$

Setting the VOM to the 2.5 VAC range and applying just enough AC voltage to bring the needle to the 0 dB mark allows us to verify that this is indeed equivalent to just under 0.8 Volts (read on the 2.5 VAC scale):



In the lower-right corner of the meter face we see some notes regarding correction values for decibel measurements when using different AC voltage ranges. The dB scale is read directly when the meter is set on the 2.5 VAC range. When set on the 10 VAC range (i.e. a range *four times* as great), the meter's needle will experience a deflection one-fourth as much as when set to the 2.5 VAC range, and therefore it will point to a lesser (or even negative) value on the dB scale. Converting a voltage ratio of 0.25 into a decibel figure shows us how much less the needle will register on the dB scale when the voltage range is quadrupled:

$$20 \log \left(\frac{2.5}{10} \right) = -12.04 \text{ dB}$$

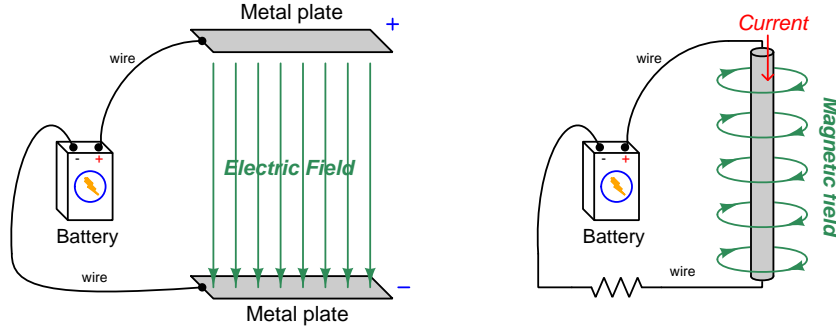
Therefore, when using the 10 VAC range instead of the 2.5 VAC range, one must add 12 dB to the reading. Likewise, we may prove each of the printed correction factors for the alternative voltage-measurement ranges listed (50 Volt AC range and 250 Volt AC range):

$$20 \log \left(\frac{2.5}{50} \right) = -26.02 \text{ dB}$$

$$20 \log \left(\frac{2.5}{250} \right) = -40.0 \text{ dB}$$

4.2 Near-field versus far-field regions

An electric field exists wherever voltage exists (i.e. wherever an imbalance exists between positive and negative electrical charges), parallel to the axis of that voltage. A magnetic field exists wherever an electric charge moves, perpendicular to the axis of that charge's motion. These phenomena exist for DC (direct) as well as AC (alternating) electricity, and are shown in the two following illustrations:



Electric and magnetic fields, however, are not simply manifestations of voltage and current, respectively. Mathematical discoveries made by the Scottish physicist James Clerk Maxwell (and later simplified by Oliver Heaviside) also relate electric fields and magnetic fields *directly to each other*. Consider two⁶ of Maxwell's equations shown below:

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_B}{dt}$$

$$\oint \vec{B} \cdot d\vec{s} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}$$

The first equation describes how an electric field (\vec{E}) is created by a varying magnetic flux ($\frac{d\Phi_B}{dt}$), which is otherwise known as *Faraday's Law of Electromagnetic Induction*. This is the law exploited in the function of electro-mechanical generators whereby coils of wire are subjected to changing magnetic fields, creating voltage between the coils' end-points. The negative sign in this equation is an embodiment of *Lenz's Law*, which states that any current resulting from the induced voltage will produce its own magnetic field opposing the first magnetic field's direction of change.

The second equation shown here describes *two* different ways to produce a magnetic field (\vec{B}). One way is to use a moving stream of electric charges known as a *current* (I), a fact also known as *Ampère's Law*. This is the law exploited in the function of electromagnets, where we produce a magnetic field by connecting a coil of wire to an electrical source. The second way is to vary an electric flux ($\frac{d\Phi_E}{dt}$) in empty space with surrounding magnetic permeability μ_0 and electric permittivity ϵ_0 .

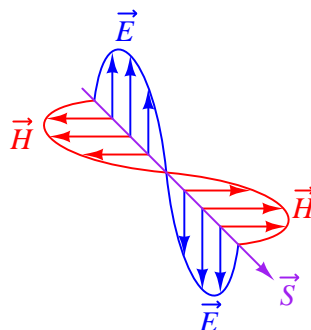
⁶Maxwell's equations of electromagnetism are typically presented as a set of *four*. Here we only list two of these, as the other two are not relevant to electromagnetic waves in particular.

Maxwell's equations suggest that in pulsed DC and AC circuits, where the resulting electric and/or magnetic fields vary over time ($\frac{d}{dt}$) rather than holding steady, those time-varying fields will mutually support one another over space and time. That is, a varying electric field will eventually produce a varying magnetic field, and vice-versa, those fields propagating one another as an *electromagnetic wave* radiating away from the circuit through space at the speed of light⁷.

At close range, immediately adjacent to the circuit conductors, electric and magnetic fields are strictly functions of those conductors' voltages and currents, respectively. These fields' shapes follow the basic forms shown in the previous illustrations: magnetic field lines circling current-carrying conductors and electric field lines stretching between conductive surfaces. We generally refer to this region of space around an energized circuit as the *near-field region*, where electric and magnetic fields are distinct from one another in form and in relative magnitude.

At farther distances from the energized circuit, the electric and magnetic fields are found in pairs oscillating at right angles to one another in space, forming a single electromagnetic wave radiating away from the source circuit spherically in all directions. This form of wave is a direct consequence of Maxwell's equations describing how the rate-of-change of one field creates the other field, and vice-versa. Below we see a crude representation of an electromagnetic wave, with an electric \vec{E} field, a magnetic \vec{H} field, and the vector of propagation \vec{S} called the *Poynting vector*:

Electromagnetic wave



This region where the electric and magnetic fields exist as coupled pairs comprising a well-formed wave is referred to as the *far-field region*. In this region the electric and magnetic fields are always perpendicular to one another and their relative magnitudes are in a fixed proportion governed by the permeability and permittivity of free space.

The demarcation between near- and far-field regions depends on the physical dimensions of the radiating circuit as well as the *wavelength* of the signal (λ), wavelength being calculated by dividing the speed of light (2.9979×10^8 meters per second in empty space) by the signal frequency in Hertz. There is no simple rule for predicting where the near-field region ends and the far-field region begins, and in fact there is a gradual morphing of one to the other, but one may reliably consider any distance from the circuit in excess of multiple wavelengths to be far-field.

⁷In fact, this is precisely what light is: an electromagnetic wave of exceptionally high frequency, well beyond the frequency range of contemporary electric circuits.

The following list compares and contrasts several near-field versus far-field characteristics:

- In the near-field region electric and magnetic field strengths depend greatly on the geometry of the energized circuit's conductors as well as on the specific voltage and current levels existing at each location. In the far-field region, however, the electric and magnetic field strengths are always in fixed proportion to one another.
- Far-field effects may be ignored for low-frequency AC circuits because these circuits' wavelengths are so long. At 60 Hz, for example, the wavelength is nearly 5 *million* meters. If we consider the far-field region of a circuit to begin at least multiple wavelengths from the source, this puts the far-field region of a 60 Hz circuit at least one-quarter of the Earth's circumference away from the circuit in question!
- In the near-field region the proportionality between electric and magnetic field strengths is a function of circuit impedance: high-impedance circuits will have stronger electric fields than magnetic fields, and low-impedance circuits will have stronger magnetic fields than electric fields, all other factors being equal. This is due to Ohm's Law establishing the relationship between voltage and current ($Z = \frac{V}{I}$), and how electric fields originate in a circuit from potential differences while magnetic fields originate in a circuit from currents. In the far-field region, however, where the electric and magnetic fields exist only through mutual support, their strengths are always in the same (fixed) proportion defined as the *characteristic impedance of free space* which is approximately 377 Ohms. This value is equal to the square root of the ratio of magnetic permeability to electric permittivity for free space ($Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}$) because these parameters of space dictate just how strongly a varying electric field creates a magnetic field.
- The relationship between field strength and physical distance from the circuit is a very complex one within the near-field region, as this depends greatly on the geometry of the circuit conductors. In the far-field region, however, we find the *Inverse-Square Law* always holds true: as distance from the circuit increases, electromagnetic wave strength simply and reliably diminishes with the *square* of that distance as if the wave were radiating away in a spherical fashion from a point-source⁸. For example, if we double the distance from the radiating circuit (from a point in the far-field region to another point in the far-field region twice as far away in empty space), the power conveyed by the electromagnetic radiation will always be *four times* less. Tripling the far-field distance always weakens the wave's power by a factor of *nine*. This is simply because the area over which the radiated energy spreads increases with the square of the distance from any point-source, viewing that point-source as the center of a sphere.
- In the far-field region every radiating circuit has a definite *radiation pattern* where the strength of the radiated electromagnetic wave as a function of the *angle* from the geometric centerline follows a consistent pattern regardless of distance. This is a very important feature of *antennas*, where different antenna designs feature unique far-field radiation patterns. By contrast, in the near-field region of an antenna the radiation pattern varies significantly with distance.

⁸A "point source" is a hypothetical point in space having no height nor width nor depth (i.e. it has *zero* physical dimensions) that emits radiation. Point-sources are a theoretical concept only. Real sources have spatial dimensions which makes their near-field strength/distance relationships complex, but the farther away you get from them the more their radiative behavior approaches that of a theoretical point-source.

- The spatial *polarization* of the electromagnetic wave (i.e. the orientation of its perpendicular electric and magnetic fields) is always well-defined in the far-field region but often ill-defined in the near-field region. This means, among other things, that two or more antennas will exchange electromagnetic energy efficiently only if they are appropriately oriented to one another over far-field distances, but may exchange energy fairly well regardless of orientation over near-field distances.
- Any “gain” specifications for an antenna structure apply only to apparent gains in signal power over the far-field range, because they refer to the degree to which an antenna focuses its energy in one direction more than another (i.e. its directionality), implying a distance over which the electromagnetic wave has become well-formed and the antenna’s radiation pattern is reliably established. At near-field distances these “gain” figures are meaningless.

4.3 The Poynting vector

We know that electric fields (E) are always associated with a voltage (V), and magnetic fields (field strength B and field force H) with a current (I). Specifically, electric field strength is the quotient of voltage and distance between two points which is why electric fields are measured in units of *Volts per meter*. Similarly, magnetic field force around a straight current-carrying conductor is the quotient of current and circumference of the circular magnetic path (equal to 2π times the distance from the conductor) which is why magnetic field force is measured in units of *Amperes per meter*.

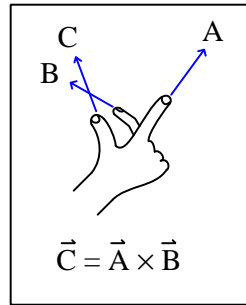
From Joule's Law we know that the rate of energy transfer (called power) whenever there is simultaneous voltage and current ($P = VI$). It stands to reason, then, that just as voltage multiplied by current yields power, an electric field multiplied by a magnetic field must similarly express power transferred by those two fields through space. That is, if $P = VI$, then the product of E and H should yield a power *intensity* which we shall label S :

$$P = VI \qquad [\text{Watts}] = [\text{Volts}][\text{Amperes}]$$

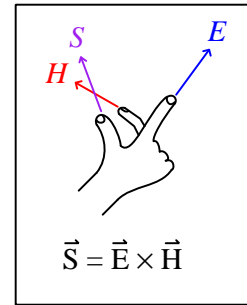
$$S = EH \qquad [\text{Watts/meter}^2] = [\text{Volts/meter}][\text{Amperes/meter}]$$

Notice how the Joule's Law equation $P = VI$ involves scalar quantities, while the $S = EH$ equation necessarily involves spatial dimensions of distance (meters) and area (square meters). This is because both electric and magnetic fields are *vector* quantities having both magnitude and spatial direction. Power intensity (S) is the measure of how much power is flowing per square meter of area in open space. As it so happens, power intensity in Watts per square meter relates to electric and magnetic fields by the right-hand rule of vector cross-products, the vector expression of that power intensity being called the *Poynting vector* (\vec{S}) after its formulator, the English physicist John H. Poynting:

General "right-hand rule"
for vector cross-products



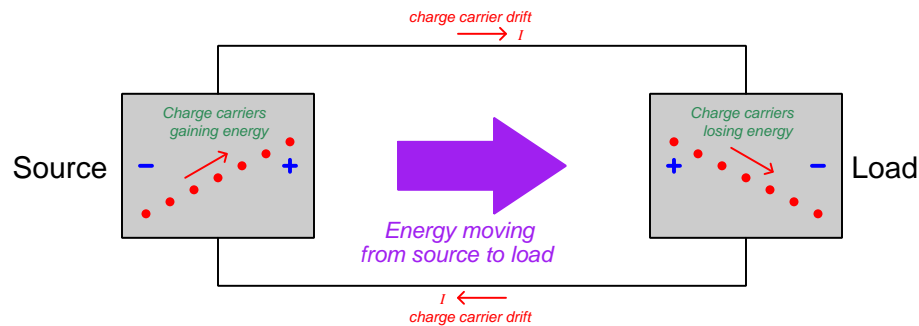
"Right-hand rule" specific
to the Poynting vector



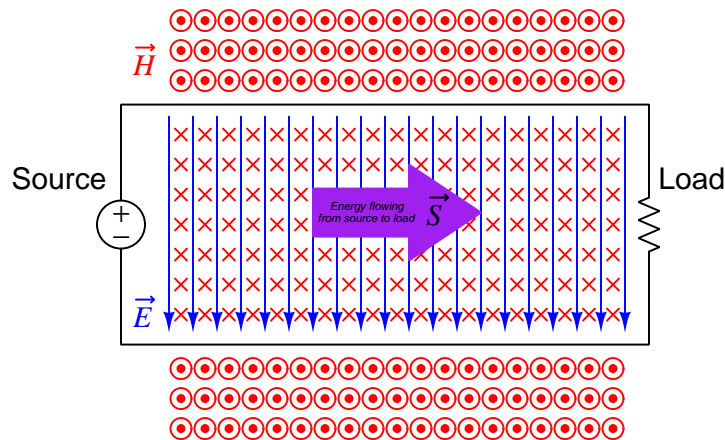
Poynting's contribution to the science of electromagnetism was expressing energy conservation and energy transfer in terms of electric and magnetic fields. That is, the electric and magnetic fields associated with any electrical *source* result in a flow of energy away from that source *through space*, and the electric and magnetic fields associated with any *load* result in a flow of energy toward that load *through space*. The direction of the Poynting vector shows the direction of that energy flow,

and the Poynting vector magnitude shows how many Watts of power flow per square meter of space.

This should be a rather startling conclusion to anyone familiar with voltage, current, and power in simple DC circuits. Beginning students of electricity regard energy transfer in simple circuits as the gain and loss of potential energy in mobile electric charges, those charges gaining energy as they pass through a source and losing energy as they pass through a load:

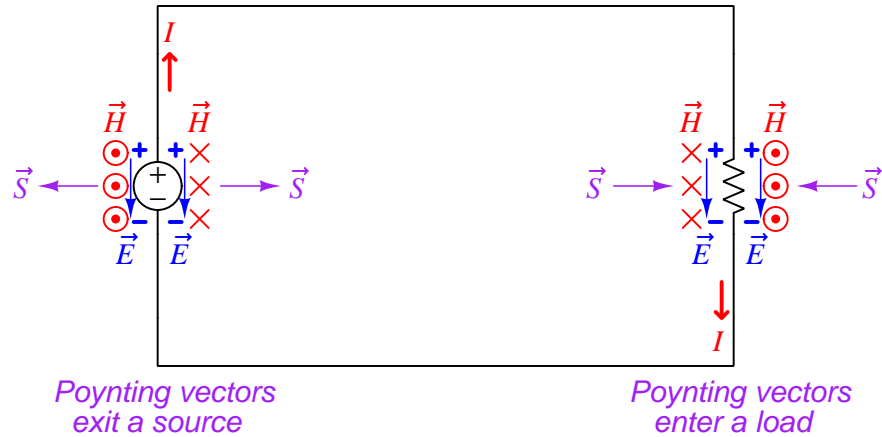


Though this model is conceptually correct, *so is Poynting's model of energy flowing through space*. In fact, if we re-draw this simple circuit illustration showing the electric and magnetic field vectors between the two wires, we can see the Poynting vector expressing that flow of energy from source to load⁹:



⁹Holding your right hand such that your index finger points downward (\vec{E}) and your middle finger into the page (\vec{H}), your thumb will point to the right (\vec{S}) from source to load.

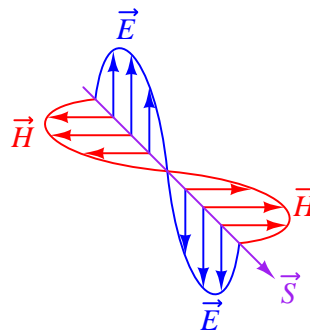
If we examine the source and load separately using the right-hand rule, we see how the Poynting vector in each case shows the proper direction of energy transfer:



It is well-worth your time to apply the right-hand rule to each of the four Poynting vectors shown above, to familiarize yourself with the use of this tool as a means to determine vector cross-product directions. For example, to the left of the source we see the electric field vector pointing down which means your index finger should point downward as well. The magnetic field vector on that side is coming out of the page (middle finger pointing toward you), and holding your right hand in that manner points your thumb to the left.

For simple DC circuits and circuit elements, Poynting's theory of power transfer may seem like a mere curiosity. It fails to explain anything we could not already explain using the model of potential energy gained and lost by mobile charge carriers, and it requires contortions of your right hand to properly align all the vectors. However, in some applications such as *radio* where energy radiates away from structures called *antennas* in the form of electromagnetic waves (oscillating electric and magnetic fields), Poynting's theory is the *only* suitable explanation for power transfer because there are no wires to convey mobile electric charges from source (transmitting antenna) to load (receiving antenna). Below we see a simplified representation of an electromagnetic wave consisting of two oscillating fields (shown here as an electric field oscillating vertically and a magnetic field oscillating horizontally), and if you use the right-hand rule in each half-cycle you will find that the Poynting vector consistently points in the same direction at the speed of light:

Electromagnetic wave



Based on the direction of the Poynting vector we see in this illustration, the source (transmitting antenna) would be located somewhere above and to the left of the drawn wave. There is no definite location for a load, as electromagnetic waves spread out in *all directions* away from the source. Our simple illustration merely shows a “cross-section” of a wave radiating *cylindrically* away from the transmitting antenna. The real electromagnetic wave (of which this illustration shows just a portion) would have an electric field resembling ripples on a pond oscillating vertically while radiating away in all horizontal directions from a single point, and have a magnetic field resembling a set of concentric circles oscillating rotationally in the horizontal plane while also radiating in all directions from that same point.

4.4 Maxwell's equations of electromagnetism

The Scottish physicist James Clerk Maxwell mathematically proved the interrelationships of electric and magnetic fields, publishing his results in the book *A Treatise on Electricity and Magnetism* in 1873. His equations were later simplified and consolidated by the work of Oliver Heaviside, commonly expressed in the following forms describing electric fields (\mathbf{E}) and magnetic fields (\mathbf{B}) within a vacuum:

$$\begin{aligned}\oint \mathbf{E} \cdot d\mathbf{A} &= \frac{Q}{\epsilon_0} \\ \oint \mathbf{B} \cdot d\mathbf{A} &= 0 \\ \oint \mathbf{E} \cdot d\mathbf{s} &= -\frac{d\Phi_B}{dt} \\ \oint \mathbf{B} \cdot d\mathbf{s} &= \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}\end{aligned}$$

We will now explore each of these four equations, relating them to laws codified by other early electrical researchers, as well as describe how they relate to the phenomenon of *electromagnetic waves*.

4.4.1 Gauss's Law of electric fields

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{Q}{\epsilon_0}$$

The first of these equations simply describes electric fields as spreading out from a quantity of electric charge (Q). If we imagine a three-dimensional surface completely enclosing an electric charge (e.g. a soap bubble surrounding a charge), the total quantity of electric flux passing through that surface will be the product of the electric field (also known as the electric flux density, \mathbf{E}) and the enclosing area (\mathbf{A}). This is also known as *Gauss's Law of electric fields*.

4.4.2 Gauss's Law of magnetic fields

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0$$

The second of these equations describes magnetic fields passing through a three-dimensional enclosing surface. If we imagine a magnet completely enclosed within a soap bubble, the sum total of magnetic flux passing out through that soap bubble's surface must equal the sum total of magnetic flux re-entering it somewhere else. Literally, this equation says the product of magnetic field (also known as magnetic flux density \mathbf{B}) and the enclosing area (\mathbf{A}) must be algebraically equal to zero. This is to say, magnetic fields must always exist as *closed loops*, or alternatively we may say there is no such thing as a magnetic monopole (i.e. a solitary North or South pole without its complement, or a *magnetic charge*). In this way magnetic fields are quite different from electric fields which radiate away from discrete electric charges without ever returning. This is also known as *Gauss's Law of magnetic fields*.

4.4.3 Faraday's Law of electromagnetic induction, extended

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$

The third equation states that an electric field (\mathbf{E}) will be produced in open space by a varying magnetic flux ($\frac{d\Phi_B}{dt}$). Michael Faraday expressed this phenomenon in terms of voltage induced along the length of a conductor by a time-varying magnetic field, but Maxwell demonstrated mathematically that this same thing happens even in empty space!

4.4.4 Ampère's Circuital Law, extended

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}$$

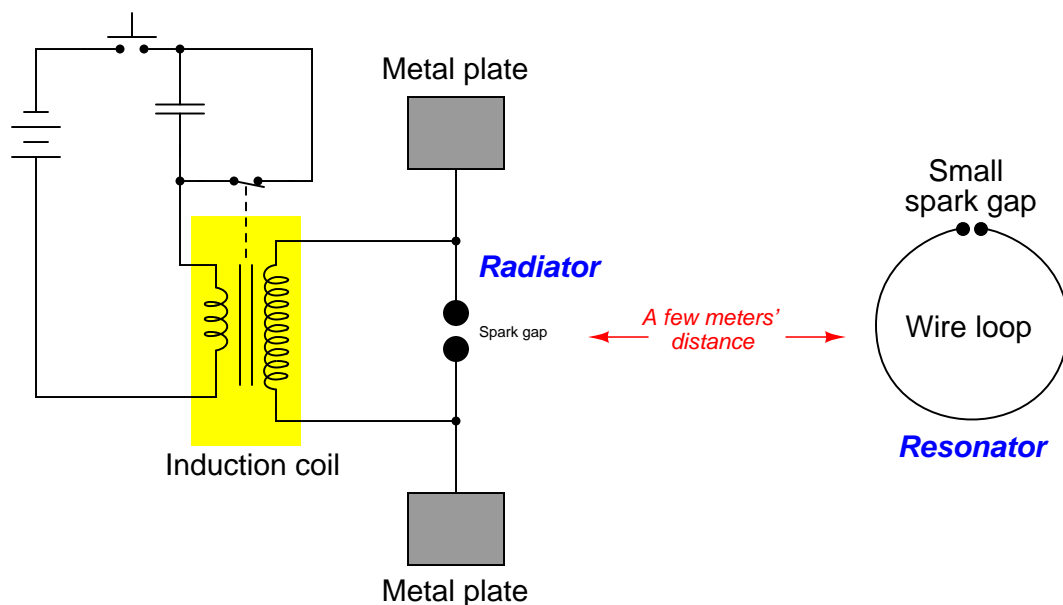
The fourth equation states that a magnetic field (\mathbf{B}) will be produced in open space either by an electric current (I) and/or by a varying electric flux ($\frac{d\Phi_E}{dt}$). The French scientist André-Marie Ampère described the relationship between current passing through a conductor and the magnetic field created around that conductor by the current, but once again Maxwell extended this concept to include magnetic fields generated in empty space. In empty space, where there are no moving electric charges (i.e. current, I) to generate a magnetic field, only the second term of the equation applies with a time-varying electric field substituting for actual electric current.

4.4.5 Proof of electromagnetic waves

Given the complementary relationship between time-varying electric and magnetic fields, Maxwell reasoned, it was possible for a varying electric field to create a varying magnetic field which would then create another varying electric field, and so on. This cause-and-effect cycle could continue, *ad infinitum*, with fast-changing electric and magnetic fields radiating off into open space without needing wires to carry or guide them, or a tank circuit to sustain the oscillations. In other words, *these complementary fields self-sustain as they travel*, although they do become weaker due to their spreading out of a finite amount of energy over larger and larger volumes of space.

The Prussian Academy of Science offered a reward to anyone who could experimentally validate Maxwell’s prediction, and this challenge was answered by Professor Heinrich Hertz at the Engineering College in Karlsruhe, Germany in 1887, eight years after Maxwell’s death. Hertz constructed and tested a pair of devices: a “radiator” to produce the electromagnetic waves, and a “resonator” to receive them.

A simplified diagram showing Hertz’s experimental device is shown here:



An “induction coil” (a buzzing device constructed of a self-interrupting relay and step-up transformer winding to generate a continuous pulsing waveform of high voltage) provided an extremely noisy (i.e. harmonically rich)¹⁰ AC signal to the radiator, while a spark gap at the

¹⁰A fundamental property of waves is that any wave-shape, no matter how complex, is mathematically equivalent to a series of perfectly sinusoidal waves of differing frequencies added together. Waveforms with fast rise/fall times, such as the pulse waveforms produced by switched induction coils, contain many such high-frequency “harmonic” signals, and it is these high-frequency signals that created the radiated electromagnetic waves in Hertz’s experiments. The induction coil’s secondary winding inductance combined with the parasitic capacitance of the radiator’s wires and plates provided the means to momentarily sustain some of these high-frequency oscillations long enough to sustain visible sparks at the resonator’s gap.

resonator provided indication that the electromagnetic waves were captured and converted into voltage by the resonator wire.

Both the radiator and the resonator are antennas. The purpose of the transmitting antenna (radiator) is to take high-frequency AC power and radiate that power in the form of electromagnetic waves: self-sustaining electric and magnetic fields propagating outward into open space. The purpose of the receiving antenna is to capture those electromagnetic waves and convert them into an electrical signal (i.e. AC voltage and current). All antennas – from historical to the most modern – behave in fundamentally the same way: energize them with high-frequency AC electrical power, and they will radiate electromagnetic waves at that frequency; expose them to electromagnetic waves, and they will produce a very small AC electrical signal at the same frequency as the incident radiation.

Chapter 5

Questions

This learning module, along with all others in the ModEL collection, is designed to be used in an inverted instructional environment where students independently read¹ the tutorials and attempt to answer questions on their own *prior* to the instructor's interaction with them. In place of lecture², the instructor engages with students in Socratic-style dialogue, probing and challenging their understanding of the subject matter through inquiry.

Answers are not provided for questions within this chapter, and this is by design. Solved problems may be found in the Tutorial and Derivation chapters, instead. The goal here is *independence*, and this requires students to be challenged in ways where others cannot think for them. Remember that you always have the tools of *experimentation* and *computer simulation* (e.g. SPICE) to explore concepts!

The following lists contain ideas for Socratic-style questions and challenges. Upon inspection, one will notice a strong theme of *metacognition* within these statements: they are designed to foster a regular habit of examining one's own thoughts as a means toward clearer thinking. As such these sample questions are useful both for instructor-led discussions as well as for self-study.

¹Technical reading is an essential academic skill for any technical practitioner to possess for the simple reason that the most comprehensive, accurate, and useful information to be found for developing technical competence is in textual form. Technical careers in general are characterized by the need for continuous learning to remain current with standards and technology, and therefore any technical practitioner who cannot read well is handicapped in their professional development. An excellent resource for educators on improving students' reading prowess through intentional effort and strategy is the book *Reading For Understanding – How Reading Apprenticeship Improves Disciplinary Learning in Secondary and College Classrooms* by Ruth Schoenbach, Cynthia Greenleaf, and Lynn Murphy.

²Lecture is popular as a teaching method because it is easy to implement: any reasonably articulate subject matter expert can talk to students, even with little preparation. However, it is also quite problematic. A good lecture always makes complicated concepts seem easier than they are, which is bad for students because it instills a false sense of confidence in their own understanding; reading and re-articulation requires more cognitive effort and serves to verify comprehension. A culture of teaching-by-lecture fosters a debilitating dependence upon direct personal instruction, whereas the challenges of modern life demand independent and critical thought made possible only by gathering information and perspectives from afar. Information presented in a lecture is ephemeral, easily lost to failures of memory and dictation; text is forever, and may be referenced at any time.

GENERAL CHALLENGES FOLLOWING TUTORIAL READING

- Summarize as much of the text as you can in one paragraph of your own words. A helpful strategy is to explain ideas as you would for an intelligent child: as simple as you can without compromising too much accuracy.
- Simplify a particular section of the text, for example a paragraph or even a single sentence, so as to capture the same fundamental idea in fewer words.
- Where did the text make the most sense to you? What was it about the text's presentation that made it clear?
- Identify where it might be easy for someone to misunderstand the text, and explain why you think it could be confusing.
- Identify any new concept(s) presented in the text, and explain in your own words.
- Identify any familiar concept(s) such as physical laws or principles applied or referenced in the text.
- Devise a proof of concept experiment demonstrating an important principle, physical law, or technical innovation represented in the text.
- Devise an experiment to disprove a plausible misconception.
- Did the text reveal any misconceptions you might have harbored? If so, describe the misconception(s) and the reason(s) why you now know them to be incorrect.
- Describe any useful problem-solving strategies applied in the text.
- Devise a question of your own to challenge a reader's comprehension of the text.

GENERAL FOLLOW-UP CHALLENGES FOR ASSIGNED PROBLEMS

- Identify where any fundamental laws or principles apply to the solution of this problem, especially before applying any mathematical techniques.
- Devise a thought experiment to explore the characteristics of the problem scenario, applying known laws and principles to mentally model its behavior.
- Describe in detail your own strategy for solving this problem. How did you identify and organized the given information? Did you sketch any diagrams to help frame the problem?
- Is there more than one way to solve this problem? Which method seems best to you?
- Show the work you did in solving this problem, even if the solution is incomplete or incorrect.
- What would you say was the most challenging part of this problem, and why was it so?
- Was any important information missing from the problem which you had to research or recall?
- Was there any extraneous information presented within this problem? If so, what was it and why did it not matter?
- Examine someone else's solution to identify where they applied fundamental laws or principles.
- Simplify the problem from its given form and show how to solve this simpler version of it. Examples include eliminating certain variables or conditions, altering values to simpler (usually whole) numbers, applying a limiting case (i.e. altering a variable to some extreme or ultimate value).
- For quantitative problems, identify the real-world meaning of all intermediate calculations: their units of measurement, where they fit into the scenario at hand. Annotate any diagrams or illustrations with these calculated values.
- For quantitative problems, try approaching it qualitatively instead, thinking in terms of “increase” and “decrease” rather than definite values.
- For qualitative problems, try approaching it quantitatively instead, proposing simple numerical values for the variables.
- Were there any assumptions you made while solving this problem? Would your solution change if one of those assumptions were altered?
- Identify where it would be easy for someone to go astray in attempting to solve this problem.
- Formulate your own problem based on what you learned solving this one.

GENERAL FOLLOW-UP CHALLENGES FOR EXPERIMENTS OR PROJECTS

- In what way(s) was this experiment or project easy to complete?
- Identify some of the challenges you faced in completing this experiment or project.

- Show how thorough documentation assisted in the completion of this experiment or project.
- Which fundamental laws or principles are key to this system's function?
- Identify any way(s) in which one might obtain false or otherwise misleading measurements from test equipment in this system.
- What will happen if (component X) fails (open/shorted/etc.)?
- What would have to occur to make this system unsafe?

5.1 Conceptual reasoning

These questions are designed to stimulate your analytic and synthetic thinking³. In a Socratic discussion with your instructor, the goal is for these questions to prompt an extended dialogue where assumptions are revealed, conclusions are tested, and understanding is sharpened. Your instructor may also pose additional questions based on those assigned, in order to further probe and refine your conceptual understanding.

Questions that follow are presented to challenge and probe your understanding of various concepts presented in the tutorial. These questions are intended to serve as a guide for the Socratic dialogue between yourself and the instructor. Your instructor's task is to ensure you have a sound grasp of these concepts, and the questions contained in this document are merely a means to this end. Your instructor may, at his or her discretion, alter or substitute questions for the benefit of tailoring the discussion to each student's needs. The only absolute requirement is that each student is challenged and assessed at a level equal to or greater than that represented by the documented questions.

It is far more important that you convey your *reasoning* than it is to simply convey a correct answer. For this reason, you should refrain from researching other information sources to answer questions. What matters here is that *you* are doing the thinking. If the answer is incorrect, your instructor will work with you to correct it through proper reasoning. A correct answer without an adequate explanation of how you derived that answer is unacceptable, as it does not aid the learning or assessment process.

You will note a conspicuous lack of answers given for these conceptual questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your conceptual answers, where applicable, is to use circuit simulation software to explore the effects of changes made to circuits. For example, if one of these conceptual questions challenges you to predict the effects of altering some component parameter in a circuit, you may check the validity of your work by simulating that same parameter change within software and seeing if the results agree.

³*Analytical* thinking involves the “disassembly” of an idea into its constituent parts, analogous to dissection. *Synthetic* thinking involves the “assembly” of a new idea comprised of multiple concepts, analogous to construction. Both activities are high-level cognitive skills, extremely important for effective problem-solving, necessitating frequent challenge and regular practice to fully develop.

5.1.1 Reading outline and reflections

“Reading maketh a full man; conference a ready man; and writing an exact man” – Francis Bacon

Francis Bacon’s advice is a blueprint for effective education: reading provides the learner with knowledge, writing focuses the learner’s thoughts, and critical dialogue equips the learner to confidently communicate and apply their learning. Independent acquisition and application of knowledge is a powerful skill, well worth the effort to cultivate. To this end, students should read these educational resources closely, journal their own reflections on the reading, and discuss in detail their findings with classmates and instructor(s). You should be able to do all of the following after reading any instructional text:

☒ Briefly SUMMARIZE THE TEXT in the form of a journal entry documenting your learning as you progress through the course of study. Share this summary in dialogue with your classmates and instructor. Journaling is an excellent self-test of thorough reading because you cannot clearly express what you have not read or did not comprehend.

☒ Demonstrate ACTIVE READING STRATEGIES, including verbalizing your impressions as you read, simplifying long passages to convey the same ideas using fewer words, annotating text and illustrations with your own interpretations, working through mathematical examples shown in the text, cross-referencing passages with relevant illustrations and/or other passages, identifying problem-solving strategies applied by the author, etc. Technical reading is a special case of problem-solving, and so these strategies work precisely because they help solve any problem: paying attention to your own thoughts (metacognition), eliminating unnecessary complexities, identifying what makes sense, paying close attention to details, drawing connections between separated facts, and noting the successful strategies of others.

☒ Identify IMPORTANT THEMES, especially GENERAL LAWS and PRINCIPLES, expounded in the text and express them in the simplest of terms as though you were teaching an intelligent child. This emphasizes connections between related topics and develops your ability to communicate complex ideas to anyone.

☒ Form YOUR OWN QUESTIONS based on the reading, and then pose them to your instructor and classmates for their consideration. Anticipate both correct and incorrect answers, the incorrect answer(s) assuming one or more plausible misconceptions. This helps you view the subject from different perspectives to grasp it more fully.

☒ Devise EXPERIMENTS to test claims presented in the reading, or to disprove misconceptions. Predict possible outcomes of these experiments, and evaluate their meanings: what result(s) would confirm, and what would constitute disproof? Running mental simulations and evaluating results is essential to scientific and diagnostic reasoning.

☒ Specifically identify any points you found CONFUSING. The reason for doing this is to help diagnose misconceptions and overcome barriers to learning.

5.1.2 Foundational concepts

Correct analysis and diagnosis of electric circuits begins with a proper understanding of some basic concepts. The following is a list of some important concepts referenced in this module's full tutorial. Define each of them in your own words, and be prepared to illustrate each of these concepts with a description of a practical example and/or a live demonstration.

Energy

Conservation of Energy

Parasitic effect

Capacitance

Inductance

Electric field

Magnetic field

Electromagnetic wave

Fundamental frequency

Harmonic frequency

Antenna

Period

Frequency

Wavelength

Directionality

Resonance

Dipole

Reciprocity

SCADA

Decibels

Parabola

Radiation pattern

Isotropic

Gain

Radiation resistance

Joule's Law

5.1.3 Parabolic dish feed

A close inspection of the following photograph reveals a small *feed antenna* element located at the focal point of the dish:



Identify the type of antenna structure used for “feeding” this dish, and explain how it functions.

Challenges

- Why is this dish antenna made of metal bars rather than being a solid reflector?

5.1.4 Antenna blind spots

Identify the “blind spots” of various antenna types – that is, physical locations near the antenna where radiation strength will be at a minimum (if the antenna is transmitting) and where the antenna will be least sensitive to any incoming signals.

- Whip =
- Dipole =
- Yagi =

Challenges

- Where is the blind spot for an isotropic antenna?

5.1.5 Dummy antenna

When radio technicians need to test the operation of a transmitter at significant power levels, but without actually broadcasting any signal through space, they disconnect the antenna from the transmitter’s RF output and in its place connect a *dummy load* consisting of a high-power resistor with a value equal to the “radiation resistance” of the normal antenna.

Explain why a resistive “dummy load” simulates an antenna realistically to the transmitter, since real antennas are simply made of metal struts and therefore contain negligible resistance.

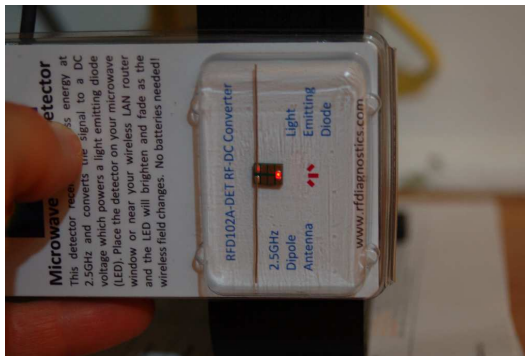
Challenges

- What practical considerations should we consider if building a dummy load for a radio transmitter of significant power, such as a 50 kW AM broadcast transmitter?

5.1.6 Microwave detector

The model RFD102A-DET microwave energy detector consists of a small dipole antenna soldered to an RF-to-DC converter circuit, which in turn is soldered to a light-emitting diode. Together, these components act as a microwave field detector efficient enough to energize the LED by the radiated power from a standard home Wi-Fi network router.

In the following photographs we see one of these detectors held near a Wi-Fi router in two orientations, one of which results in the LED illuminating and the other resulting in no light:



Explain why the orientation matters to the LED's illumination.

Challenges

- Identify one additional orientation we should test the detector in, to make this experiment thorough.

5.1.7 Transmitter testing

The Federal Communication Commission (FCC) bulletin number 63 (“Understanding the FCC Regulations for Low-Power, Non-Licensed Transmitters”) contains the following paragraph:

A particular transmitter that generates a constant level of power (Watts) can produce electric fields of different strengths ($\mu\text{V}/\text{m}$) depending on, among other things, the type of transmission line and antenna connected to it. Because it is the electric field that causes interference to authorized radio communications, and since a particular electric field strength does not directly correspond to a particular level of transmitter power, most of the Part 15 emission limits are specified in field strength.

Explain why measured electric field strength doesn’t necessarily correspond to transmitter output power. What are some of the reasons why these two may not be proportional? Be as specific as you can in your answers.

Challenges

- What shape of antenna might be best for intercepting the electric field component of an electromagnetic wave, versus intercepting the magnetic field component?

5.1.8 Close-coupled antennas

Suppose two high-gain antennas were placed extremely close to each other, so that the distance separating them was negligible. Would the power received at the receiving antenna exceed the power radiated by the transmitting antenna? Why or why not?

Challenges

- What should be the maximum antenna-to-antenna gain achievable in this configuration?

5.2 Quantitative reasoning

These questions are designed to stimulate your computational thinking. In a Socratic discussion with your instructor, the goal is for these questions to reveal your mathematical approach(es) to problem-solving so that good technique and sound reasoning may be reinforced. Your instructor may also pose additional questions based on those assigned, in order to observe your problem-solving firsthand.

Mental arithmetic and estimations are strongly encouraged for all calculations, because without these abilities you will be unable to readily detect errors caused by calculator misuse (e.g. keystroke errors).

You will note a conspicuous lack of answers given for these quantitative questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. My advice is to use circuit simulation software such as SPICE to check the correctness of quantitative answers. Refer to those learning modules within this collection focusing on SPICE to see worked examples which you may use directly as practice problems for your own study, and/or as templates you may modify to run your own analyses and generate your own practice problems.

Completely worked example problems found in the Tutorial may also serve as “test cases⁴” for gaining proficiency in the use of circuit simulation software, and then once that proficiency is gained you will never need to rely⁵ on an answer key!

⁴In other words, set up the circuit simulation software to analyze the same circuit examples found in the Tutorial. If the simulated results match the answers shown in the Tutorial, it confirms the simulation has properly run. If the simulated results disagree with the Tutorial’s answers, something has been set up incorrectly in the simulation software. Using every Tutorial as practice in this way will quickly develop proficiency in the use of circuit simulation software.

⁵This approach is perfectly in keeping with the instructional philosophy of these learning modules: *teaching students to be self-sufficient thinkers*. Answer keys can be useful, but it is even more useful to your long-term success to have a set of tools on hand for checking your own work, because once you have left school and are on your own, there will no longer be “answer keys” available for the problems you will have to solve.

5.2.1 Miscellaneous physical constants

Note: constants shown in **bold** type are *exact*, not approximations. Values inside of parentheses show one standard deviation (σ) of uncertainty in the final digits: for example, the magnetic permeability of free space value given as $1.25663706212(19) \times 10^{-6}$ H/m represents a center value (i.e. the location parameter) of $1.25663706212 \times 10^{-6}$ Henrys per meter with one standard deviation of uncertainty equal to $0.0000000000019 \times 10^{-6}$ Henrys per meter.

Avogadro's number (N_A) = **$6.02214076 \times 10^{23}$** per mole (mol^{-1})

Boltzmann's constant (k) = **1.380649×10^{-23}** Joules per Kelvin (J/K)

Electronic charge (e) = **$1.602176634 \times 10^{-19}$** Coulomb (C)

Faraday constant (F) = **$96,485.33212...$** $\times 10^4$ Coulombs per mole (C/mol)

Magnetic permeability of free space (μ_0) = $1.25663706212(19) \times 10^{-6}$ Henrys per meter (H/m)

Electric permittivity of free space (ϵ_0) = $8.8541878128(13) \times 10^{-12}$ Farads per meter (F/m)

Characteristic impedance of free space (Z_0) = $376.730313668(57)$ Ohms (Ω)

Gravitational constant (G) = $6.67430(15) \times 10^{-11}$ cubic meters per kilogram-seconds squared ($\text{m}^3/\text{kg}\cdot\text{s}^2$)

Molar gas constant (R) = **$8.314462618...$** Joules per mole-Kelvin (J/mol-K) = $0.08205746(14)$ liters-atmospheres per mole-Kelvin

Planck constant (h) = **$6.62607015 \times 10^{-34}$** joule-seconds (J-s)

Stefan-Boltzmann constant (σ) = **$5.670374419...$** $\times 10^{-8}$ Watts per square meter-Kelvin⁴ ($\text{W}/\text{m}^2\cdot\text{K}^4$)

Speed of light in a vacuum (c) = **$299,792,458$** meters per second (m/s) = 186282.4 miles per second (mi/s)

Note: All constants taken from NIST data "Fundamental Physical Constants – Complete Listing", from <http://physics.nist.gov/constants>, National Institute of Standards and Technology (NIST), 2018 CODATA Adjustment.

5.2.2 Introduction to spreadsheets

A powerful computational tool you are encouraged to use in your work is a *spreadsheet*. Available on most personal computers (e.g. Microsoft Excel), *spreadsheet* software performs numerical calculations based on number values and formulae entered into cells of a grid. This grid is typically arranged as lettered columns and numbered rows, with each cell of the grid identified by its column/row coordinates (e.g. cell B3, cell A8). Each cell may contain a string of text, a number value, or a mathematical formula. The spreadsheet automatically updates the results of all mathematical formulae whenever the entered number values are changed. This means it is possible to set up a spreadsheet to perform a series of calculations on entered data, and those calculations will be re-done by the computer any time the data points are edited in any way.

For example, the following spreadsheet calculates average speed based on entered values of distance traveled and time elapsed:

	A	B	C	D
1	Distance traveled	46.9	Kilometers	
2	Time elapsed	1.18	Hours	
3	Average speed	= B1 / B2	km/h	
4				
5				

Text labels contained in cells A1 through A3 and cells C1 through C3 exist solely for readability and are not involved in any calculations. Cell B1 contains a sample distance value while cell B2 contains a sample time value. The formula for computing speed is contained in cell B3. Note how this formula begins with an “equals” symbol (=), references the values for distance and speed by lettered column and numbered row coordinates (B1 and B2), and uses a forward slash symbol for division (/). The coordinates B1 and B2 function as *variables*⁶ would in an algebraic formula.

When this spreadsheet is executed, the numerical value 39.74576 will appear in cell B3 rather than the formula = B1 / B2, because 39.74576 is the computed speed value given 46.9 kilometers traveled over a period of 1.18 hours. If a different numerical value for distance is entered into cell B1 or a different value for time is entered into cell B2, cell B3’s value will automatically update. All you need to do is set up the given values and any formulae into the spreadsheet, and the computer will do all the calculations for you.

Cell B3 may be referenced by other formulae in the spreadsheet if desired, since it is a variable just like the given values contained in B1 and B2. This means it is possible to set up an entire chain of calculations, one dependent on the result of another, in order to arrive at a final value. The arrangement of the given data and formulae need not follow any pattern on the grid, which means you may place them anywhere.

⁶Spreadsheets may also provide means to attach text labels to cells for use as variable names (Microsoft Excel simply calls these labels “names”), but for simple spreadsheets such as those shown here it’s usually easier just to use the standard coordinate naming for each cell.

Common⁷ arithmetic operations available for your use in a spreadsheet include the following:

- Addition (+)
- Subtraction (-)
- Multiplication (*)
- Division (/)
- Powers (^)
- Square roots (sqrt())
- Logarithms (ln() , log10())

Parentheses may be used to ensure⁸ proper order of operations within a complex formula. Consider this example of a spreadsheet implementing the *quadratic formula*, used to solve for roots of a polynomial expression in the form of $ax^2 + bx + c$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

	A	B
1	x_1	= (-B4 + sqrt((B4^2) - (4*B3*B5))) / (2*B3)
2	x_2	= (-B4 - sqrt((B4^2) - (4*B3*B5))) / (2*B3)
3	a =	9
4	b =	5
5	c =	-2

This example is configured to compute roots⁹ of the polynomial $9x^2 + 5x - 2$ because the values of 9, 5, and -2 have been inserted into cells B3, B4, and B5, respectively. Once this spreadsheet has been built, though, it may be used to calculate the roots of *any* second-degree polynomial expression simply by entering the new a , b , and c coefficients into cells B3 through B5. The numerical values appearing in cells B1 and B2 will be automatically updated by the computer immediately following any changes made to the coefficients.

⁷Modern spreadsheet software offers a bewildering array of mathematical functions you may use in your computations. I recommend you consult the documentation for your particular spreadsheet for information on operations other than those listed here.

⁸Spreadsheet programs, like text-based programming languages, are designed to follow standard order of operations by default. However, my personal preference is to use parentheses even where strictly unnecessary just to make it clear to any other person viewing the formula what the intended order of operations is.

⁹Reviewing some algebra here, a *root* is a value for x that yields an overall value of zero for the polynomial. For this polynomial ($9x^2 + 5x - 2$) the two roots happen to be $x = 0.269381$ and $x = -0.82494$, with these values displayed in cells B1 and B2, respectively upon execution of the spreadsheet.

Alternatively, one could break up the long quadratic formula into smaller pieces like this:

$$y = \sqrt{b^2 - 4ac} \quad z = 2a$$

$$x = \frac{-b \pm y}{z}$$

	A	B	C
1	x_1	= (-B4 + C1) / C2	= sqrt ((B4^2) - (4*B3*B5))
2	x_2	= (-B4 - C1) / C2	= 2*B3
3	a =	9	
4	b =	5	
5	c =	-2	

Note how the square-root term (y) is calculated in cell C1, and the denominator term (z) in cell C2. This makes the two final formulae (in cells B1 and B2) simpler to interpret. The positioning of all these cells on the grid is completely arbitrary¹⁰ – all that matters is that they properly reference each other in the formulae.

Spreadsheets are particularly useful for situations where the same set of calculations representing a circuit or other system must be repeated for different initial conditions. The power of a spreadsheet is that it automates what would otherwise be a tedious set of calculations. One specific application of this is to simulate the effects of various components within a circuit failing with abnormal values (e.g. a shorted resistor simulated by making its value nearly zero; an open resistor simulated by making its value extremely large). Another application is analyzing the behavior of a circuit design given new components that are out of specification, and/or aging components experiencing drift over time.

¹⁰My personal preference is to locate all the “given” data in the upper-left cells of the spreadsheet grid (each data point flanked by a sensible name in the cell to the left and units of measurement in the cell to the right as illustrated in the first distance/time spreadsheet example), sometimes coloring them in order to clearly distinguish which cells contain entered data versus which cells contain computed results from formulae. I like to place all formulae in cells below the given data, and try to arrange them in logical order so that anyone examining my spreadsheet will be able to figure out *how* I constructed a solution. This is a general principle I believe all computer programmers should follow: *document and arrange your code to make it easy for other people to learn from it.*

5.2.3 Antenna dimensions

A dipole antenna designed to resonate at a fundamental frequency of 1.8 GHz will have a total length of approximately 0.083275 meters, based on the fact that dipole antennas operate at half the wavelength ($\frac{\lambda}{2}$) of the RF energy transmitted and/or received.

If we wished to calculate this same length in a different unit of measurement such as *centimeters*, we could do so using a simple fraction-based technique whereby we set up the original quantity as a fraction and then multiply by other fractions consisting of *physically equal* numerator and denominator quantities such that the old unit is replaced by a new unit:

$$\left(\frac{0.083275 \text{ m}}{1}\right) \left(\frac{100 \text{ cm}}{1 \text{ m}}\right) = 8.3275 \text{ cm}$$

We may continue with this same unit-conversion strategy to compute an answer in inches, knowing there are exactly 2.54 centimeters to one inch:

$$\left(\frac{0.083275 \text{ m}}{1}\right) \left(\frac{100 \text{ cm}}{1 \text{ m}}\right) \left(\frac{1 \text{ in}}{2.54 \text{ cm}}\right) = 3.2785 \text{ in}$$

. . . and again, to compute an answer in feet:

$$\left(\frac{0.083275 \text{ m}}{1}\right) \left(\frac{100 \text{ cm}}{1 \text{ m}}\right) \left(\frac{1 \text{ in}}{2.54 \text{ cm}}\right) \left(\frac{1 \text{ ft}}{12 \text{ in}}\right) = 0.27321 \text{ ft}$$

This method of unit-conversion offers several advantages over other methods. First, one need only commit a few conversion ratios to memory in order to convert back and forth between many combinations of units rather than memorize (or research) conversion ratios for every combination of units encountered. Second, the accuracy of the answer is assured by careful inspection of the canceled units, which is important because this tends to be the most common way in which unit-conversions go astray: by accidentally dividing when multiplication is necessary, or vice-versa.

Knowing this, determine the dimensions of a “whip” antenna designed to resonate at the fundamental frequency of 840 MHz, expressing your answer in units of meters, centimeters, feet, and inches.

Both whip and dipole antennas operate effectively at odd harmonics as well as the fundamental. Based on this fact, calculate some other operating frequencies that would be a good match for this same antenna.

Challenges

- Convert 40 miles into kilometers using nothing but the following ratios: 2.54 cm to 1 inch, 5280 feet to 1 mile, 12 inches to 1 foot, and of course metric prefixes (centi = 100, kilo = 1000, etc.).

5.2.4 Transmitter power and field strength

A simple pair of formulae relating approximate transmitted power to measured electric field and magnetic field strengths, respectively, useful when estimating the RF power output by a radio transmitter at a certain distance away based on measurements of the E field, is as follows:

$$S = \frac{PG}{4\pi D^2} = \frac{E^2}{Z_0} = H^2 Z_0$$

Where,

S = Power density of field (the Poynting vector), in Watts per square meter

P = Transmitter output power, in Watts

G = Antenna gain (1 = theoretical “isotropic” antenna)

D = Distance between transmitter antenna and receiving instrument

E = Electric field strength detected by receiving instrument, in Volts per meter

H = Magnetic field strength detected by receiving instrument, in Amperes per meter

Z_0 = Characteristic impedance of free space, approximately 376.73 Ohms

The $4\pi D^2$ term is a calculation of spherical surface area, used to estimate how much the emitted RF energy “spreads out” from the antenna based on the assumption that it radiates away equally in all directions.

First, algebraically simplify this formula as much as possible to eliminate any unnecessary constant values.

Next, algebraically manipulate this formula to solve for P given a measured value E .

Next, replace G in that manipulated formula with the antenna’s gain figure in dBi .

Lastly, estimate the following transmitter powers given these measured electric field strength (E) values and distances:

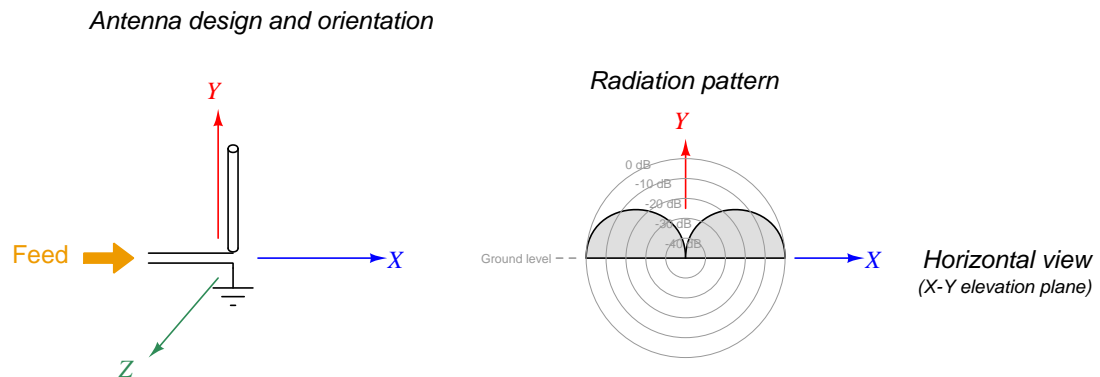
- $E = 170$ microVolts per meter ; $D = 40$ meters ; Antenna gain = +2 dBi ; $P =$
- $E = 800$ microVolts per meter ; $D = 15$ meters ; Antenna gain = +6 dBi ; $P =$
- $E = 19$ milliVolts per meter ; $D = 55$ meters ; Antenna gain = +3 dBi ; $P =$
- $E = 0.25$ Volts per meter ; $D = 8$ meters ; Antenna gain = +4 dBi ; $P =$
- $E = 1.3$ Volts per meter ; $D = 22$ meters ; Antenna gain = +10 dBi ; $P =$

Challenges

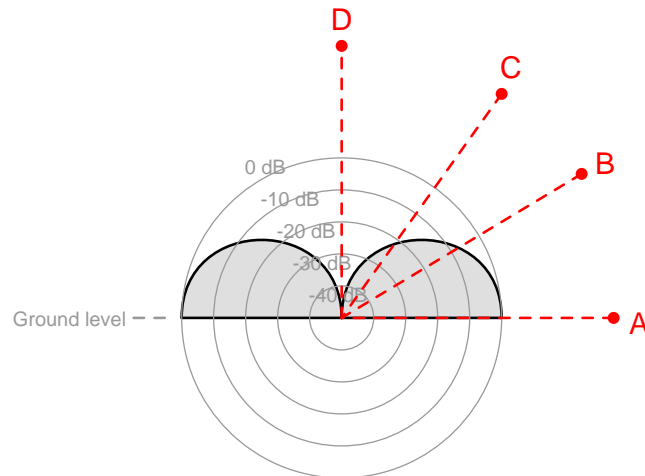
- If you double the distance between the transmitting antenna and the receiving instrument, what will this do to the measured field strength?
- These formula assume certain conditions, namely that the receiver is in the *far-field* region of the transmitting antenna (i.e. not too close), that there are no impedance mismatches in the receiver's antenna system, and that the receiving antenna is properly oriented. Explain why each of these caveats matter.

5.2.5 Monopole radiation strength

A particular monopole, or “whip”, antenna has a radiation pattern as shown below:



Identify the approximate amount of signal attenuation experienced as a result of receiving this antenna's transmitted radiation at different elevations (i.e. different angles with respect to Earth ground and the antenna's location):



- Receiver A =
- Receiver B =
- Receiver C =
- Receiver D =

Challenges

- Antenna radiation pattern plots are an example of a *polar* plot. Explain how this type of mathematical plot differs from the more conventional *Cartesian* style.

5.3 Diagnostic reasoning

These questions are designed to stimulate your deductive and inductive thinking, where you must apply general principles to specific scenarios (deductive) and also derive conclusions about the failed circuit from specific details (inductive). In a Socratic discussion with your instructor, the goal is for these questions to reinforce your recall and use of general circuit principles and also challenge your ability to integrate multiple symptoms into a sensible explanation of what's wrong in a circuit. Your instructor may also pose additional questions based on those assigned, in order to further challenge and sharpen your diagnostic abilities.

As always, your goal is to fully *explain* your analysis of each problem. Simply obtaining a correct answer is not good enough – you must also demonstrate sound reasoning in order to successfully complete the assignment. Your instructor's responsibility is to probe and challenge your understanding of the relevant principles and analytical processes in order to ensure you have a strong foundation upon which to build further understanding.

You will note a conspicuous lack of answers given for these diagnostic questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your diagnostic answers, where applicable, is to use circuit simulation software to explore the effects of faults placed in circuits. For example, if one of these diagnostic questions requires that you predict the effect of an open or a short in a circuit, you may check the validity of your work by simulating that same fault (substituting a very high resistance in place of that component for an open, and substituting a very low resistance for a short) within software and seeing if the results agree.

5.3.1 Whip antenna orientations

An RF consultant is summoned onto an industrial facility to diagnose a problem the technicians there are experiencing with some newly-installed wireless sensors. Each sensor has a “whip” style antenna on it with a swivel base allowing it to be positioned at any angle. When the consultant arrives, she finds each and every one of these antennas pointed straight at the base station (MTU) for the sensor network.

Identify the problem, and explain how to remedy it.

Challenges

- How well would the system work if all the whip antennas were angled so as to be parallel with the ground.

5.3.2 Incorrect decibel calculation program

An important expression of power ratio used in radio systems and antenna design is the *decibel*. Simply defined, “decibels” are the common logarithm of a power ratio, as shown by this formula:

$$\text{dB} = 10 \log \left(\frac{P_{out}}{P_{in}} \right)$$

If we take the common logarithm of a power ratio and then multiply that result by ten, we’ll have that power ratio’s value in decibels.

The following computer program (written in the C language) is supposed to take a user-entered value for power ratio and convert that into a decibel value. However, the program as written has an error. The source code compiles without any warnings or errors, but when executed we find the calculated decibel values are not correct:

```
#include <stdio.h>
#include <math.h>

int main (void)
{
    float pr;

    printf("Enter power ratio (unitless): ");
    scanf("%f", &pr);

    printf("A ratio of %f is equivalent to %f dB\n", pr, 10 * log(pr));

    return 0;
}
```

Identify the error and then fix it.

Challenges

- Identify good “test values” you could enter into this program to check for errors.

Appendix A

Problem-Solving Strategies

The ability to solve complex problems is arguably one of the most valuable skills one can possess, and this skill is particularly important in any science-based discipline.

- Study principles, not procedures. Don't be satisfied with merely knowing how to compute solutions – learn *why* those solutions work.
- Identify what it is you need to solve, identify all relevant data, identify all units of measurement, identify any general principles or formulae linking the given information to the solution, and then identify any “missing pieces” to a solution. Annotate all diagrams with this data.
- Sketch a diagram to help visualize the problem. When building a real system, always devise a plan for that system and analyze its function *before* constructing it.
- Follow the units of measurement and meaning of every calculation. If you are ever performing mathematical calculations as part of a problem-solving procedure, and you find yourself unable to apply each and every intermediate result to some aspect of the problem, it means you don't understand what you are doing. Properly done, every mathematical result should have practical meaning for the problem, and not just be an abstract number. You should be able to identify the proper units of measurement for each and every calculated result, and show where that result fits into the problem.
- Perform “thought experiments” to explore the effects of different conditions for theoretical problems. When troubleshooting real systems, perform *diagnostic tests* rather than visually inspecting for faults, the best diagnostic test being the one giving you the most information about the nature and/or location of the fault with the fewest steps.
- Simplify the problem until the solution becomes obvious, and then use that obvious case as a model to follow in solving the more complex version of the problem.
- Check for exceptions to see if your solution is incorrect or incomplete. A good solution will work for *all* known conditions and criteria. A good example of this is the process of testing scientific hypotheses: the task of a scientist is not to find support for a new idea, but rather to *challenge* that new idea to see if it holds up under a battery of tests. The philosophical

principle of *reductio ad absurdum* (i.e. disproving a general idea by finding a specific case where it fails) is useful here.

- Work “backward” from a hypothetical solution to a new set of given conditions.
- Add quantities to problems that are qualitative in nature, because sometimes a little math helps illuminate the scenario.
- Sketch graphs illustrating how variables relate to each other. These may be quantitative (i.e. with realistic number values) or qualitative (i.e. simply showing increases and decreases).
- Treat quantitative problems as qualitative in order to discern the relative magnitudes and/or directions of change of the relevant variables. For example, try determining what happens if a certain variable were to increase or decrease before attempting to precisely calculate quantities: how will each of the dependent variables respond, by increasing, decreasing, or remaining the same as before?
- Consider limiting cases. This works especially well for qualitative problems where you need to determine which direction a variable will change. Take the given condition and magnify that condition to an extreme degree as a way of simplifying the direction of the system’s response.
- Check your work. This means regularly testing your conclusions to see if they make sense. This does *not* mean repeating the same steps originally used to obtain the conclusion(s), but rather to use some other means to check validity. Simply repeating procedures often leads to *repeating the same errors* if any were made, which is why alternative paths are better.

Appendix B

Instructional philosophy

B.1 First principles of learning

- **Anyone can learn anything** given appropriate time, effort, resources, challenges, encouragement, and expectations. Dedicating time and investing effort are the student's responsibility; providing resources, challenges, and encouragement are the teacher's responsibility; high expectations are a responsibility shared by both student and teacher.
- **Transfer is not automatic.** The human mind has a natural tendency to compartmentalize information, which means the process of taking knowledge learned in one context and applying it to another usually does not come easy and therefore should never be taken for granted.
- **Learning is iterative.** The human mind rarely learns anything perfectly on the first attempt. Anticipate mistakes and plan for multiple tries to achieve full understanding, using the lessons of those mistakes as feedback to guide future attempts.
- **Information is absorbed, but understanding is created.** Facts and procedures may be memorized easily enough by repeated exposure, but the ability to reliably apply principles to novel scenarios only comes through intense personal effort. This effort is fundamentally creative in nature: explaining new concepts in one's own words, running experiments to test understanding, building projects, and teaching others are just a few ways to creatively apply new knowledge. These acts of making knowledge "one's own" need not be perfect in order to be effective, as the value lies in the activity and not necessarily the finished product.
- **Education trumps training.** There is no such thing as an entirely isolated subject, as all fields of knowledge are connected. Training is narrowly-focused and task-oriented. Education is broad-based and principle-oriented. When preparing for a life-long career, education beats training every time.
- **Character matters.** Poor habits are more destructive than deficits of knowledge or skill. This is especially true in collective endeavors, where a team's ability to function depends on trust between its members. Simply put, no one wants an untrustworthy person on their team. An essential component of education then, is character development.
- **People learn to be responsible by bearing responsibility.** An irresponsible person is someone who has never *had* to be responsible for anything that mattered enough to them. Just as anyone can learn anything, anyone can become responsible if the personal cost of irresponsibility becomes high enough.
- **What gets measured, gets done.** Accurate and relevant assessment of learning is key to ensuring all students learn. Therefore, it is imperative to measure what matters.
- **Failure is nothing to fear.** Every human being fails, and fails in multiple ways at multiple times. Eventual success only happens when we don't stop trying.

B.2 Proven strategies for instructors

- Assume every student is capable of learning anything they desire given the proper conditions. Treat them as capable adults by granting real responsibility and avoiding artificial incentives such as merit or demerit points.
- Create a consistent culture of high expectations across the entire program of study. Demonstrate and encourage patience, persistence, and a healthy sense of self-skepticism. Anticipate and de-stigmatize error. Teach respect for the capabilities of others as well as respect for one's own fallibility.
- Replace lecture with “inverted” instruction, where students first encounter new concepts through reading and then spend class time in Socratic dialogue with the instructor exploring those concepts and solving problems individually. There is a world of difference between observing someone solve a problem versus actually solving a problem yourself, and so the point of this form of instruction is to place students in a position where they *cannot* passively observe.
- Require students to read extensively, write about what they learn, and dialogue with you and their peers to sharpen their understanding. Apply Francis Bacon's advice that “reading maketh a full man; conference a ready man; and writing an exact man”. These are complementary activities helping students expand their confidence and abilities.
- Use artificial intelligence (AI) to challenge student understanding rather than merely provide information. Find productive ways for AI to critique students' clarity of thought and of expression, for example by employing AI as a Socratic-style interlocutor or as a reviewer of students' journals. Properly applied, AI has the ability to expand student access to critical review well outside the bounds of their instructor's reach.
- Build frequent and rapid feedback into the learning process so that students know at all times how well they are learning, to identify problems early and fix them before they grow. Model the intellectual habit of self-assessing and self-correcting your own understanding (i.e. a cognitive *feedback loop*), encouraging students to do the same.
- Use “mastery” as the standard for every assessment, which means the exam or experiment or project must be done with 100% competence in order to pass. Provide students with multiple opportunity for re-tries (different versions of the assessment every time).
- Require students to devise their own hypotheses and procedures on all experiments, so that the process is truly a scientific one. Have students assess their proposed experimental procedures for risk and devise mitigations for those risks. Let nothing be pre-designed about students' experiments other than a stated task (i.e. what principle the experiment shall test) at the start and a set of demonstrable knowledge and skill objectives at the end.
- Have students build as much of their lab equipment as possible: building power sources, building test assemblies¹, and building complete working systems (no kits!). In order to provide

¹In the program I teach, every student builds their own “Development Board” consisting of a metal chassis with DIN rail, terminal blocks, and an AC-DC power supply of their own making which functions as a portable lab environment they can use at school as well as take home.

this same “ground-up” experience for every new student, this means either previous students take their creations with them, or the systems get disassembled in preparation for the new students, or the systems grow and evolve with each new student group.

- Incorporate external accountability for you and for your students, continuously improving the curriculum and your instructional methods based on proven results. Have students regularly network with active professionals through participation in advisory committee meetings, service projects, tours, jobshadows, internships, etc. Practical suggestions include requiring students to design and build projects for external clients (e.g. community groups, businesses, different departments within the institution), and also requiring students attend all technical advisory committee meetings and dialogue with the industry representatives attending.
- Repeatedly explore difficult-to-learn concepts across multiple courses, so that students have multiple opportunities to build their understanding.
- Relate all new concepts, whenever possible, to previous concepts and to relevant physical laws. Challenge each and every student, every day, to *reason* from concept to concept and to explain the logical connections between. Challenge students to verify their conclusions by multiple approaches (e.g. double-checking their work using different methods). Ask “*Why?*” often.
- Maintain detailed records on each student’s performance and share these records privately with them. These records should include academic performance as well as professionally relevant behavioral tendencies.
- Address problems while they are small, before they grow larger. This is equally true when helping students overcome confusion as it is when helping students build professional habits.
- Build rigorous quality control into the curriculum to ensure every student masters every important concept, and that the mastery is retained over time. This includes (1) review questions added to every exam to re-assess knowledge taught in previous terms, (2) cumulative exams at the end of every term to re-assess all important concepts back to the very beginning of the program, and (3) review assessments in practical (hands-on) coursework to ensure critically-important skills were indeed taught and are still retained. What you will find by doing this is that it actually boosts retention of students by ensuring that important knowledge gets taught and is retained over long spans of time. In the absence of such quality control, student learning and retention tends to be spotty and this contributes to drop-out and failure rates later in their education.
- Finally, *never rush learning*. Education is not a race. Give your students ample time to digest complex ideas, as you continually remind yourself of just how long it took you to achieve mastery! Long-term retention and the consistently correct application of concepts are always the result of *focused effort over long periods of time* which means there are no shortcuts to learning.

B.3 Proven strategies for students

The single most important piece of advice I have for any student of any subject is to take responsibility for your own development in all areas of life including mental development. Expecting others in your life to entirely guide your own development is a recipe for disappointment. This is just as true for students enrolled in formal learning institutions as it is for auto-didacts pursuing learning entirely on their own. Learning to think in new ways is key to being able to gainfully use information, to make informed decisions about your life, and to best serve those you care about. With this in mind, I offer the following advice to students:

- **Approach all learning as valuable.** No matter what course you take, no matter who you learn from, no matter the subject, there is something useful in every learning experience. If you don't see the value of every new experience, you are not looking closely enough!
- **Continually challenge yourself.** Let other people take shortcuts and find easy answers to easy problems. The purpose of education is to stretch your mind, in order to shape it into a more powerful tool. This doesn't come by taking the path of least resistance. An excellent analogy for an empowering education is productive physical exercise: becoming stronger, more flexible, and more persistent only comes through intense personal effort.
- **Master the use of language.** This includes reading extensively, writing every day, listening closely, and speaking articulately. To a great extent language channels and empowers thought, so the better you are at wielding language the better you will be at grasping abstract concepts and articulating them not only for your benefit but for others as well.
- **Do not limit yourself to the resources given to you.** Read books that are not on the reading list. Run experiments that aren't assigned to you. Form study groups outside of class. Take an entrepreneurial approach to your own education, as though it were a business you were building for your future benefit.
- **Express and share what you learn.** Take every opportunity to teach what you have learned to others, as this will not only help them but will also strengthen your own understanding².
- Realize that **no one can give you understanding**, just as no one can give you physical fitness. These both must be *built*.
- **Above all, recognize that learning is hard work, and that a certain level of frustration is unavoidable.** There are times when you will struggle to grasp some of these concepts, and that struggle is a natural thing. Take heart that it will yield with persistent and varied³ effort, and never give up! That concepts don't immediately come to you is not a sign of something wrong, but rather of something right: that you have found a worthy challenge!

²On a personal note, I was surprised to learn just how much my own understanding of electronics and related subjects was strengthened by becoming a teacher. When you are tasked every day with helping other people grasp complex topics, it catalyzes your own learning by giving you powerful incentives to study, to articulate your thoughts, and to reflect deeply on the process of learning.

³As the old saying goes, "Insanity is trying the same thing over and over again, expecting different results." If you find yourself stumped by something in the text, you should attempt a different approach. Alter the thought experiment, change the mathematical parameters, do whatever you can to see the problem in a slightly different light, and then the solution will often present itself more readily.

B.4 Design of these learning modules

“The unexamined circuit is not worth energizing” – Socrates (if he had taught electricity)

These learning modules, although useful for self-study, were designed to be used in a formal learning environment where a subject-matter expert challenges students to digest the content and exercise their critical thinking abilities in the answering of questions and in the construction and testing of working circuits. Every effort has been made to embed the following instructional and assessment philosophies within:

- The first goal of education is to enhance clear and independent thought, in order that every student reach their fullest potential in a highly complex and inter-dependent world. Robust reasoning is *always* more important than particulars of any subject matter, because its application is universal.
- Literacy is fundamental to independent learning and thought because text continues to be the most efficient way to communicate complex ideas over space and time. Those who cannot read with ease are limited in their ability to acquire knowledge and perspective.
- Articulate communication is fundamental to work that is complex and interdisciplinary.
- Faulty assumptions and poor reasoning are best corrected through challenge, not presentation. The rhetorical technique of *reductio ad absurdum* (disproving an assertion by exposing an absurdity) works well to discipline student’s minds, not only to correct the problem at hand but also to learn how to detect and correct future errors.
- Important principles should be repeatedly explored and widely applied throughout a course of study, not only to reinforce their importance and help ensure their mastery, but also to showcase the interconnectedness and utility of knowledge.

These learning modules were expressly designed to be used in an “inverted” teaching environment⁴ where students first read the introductory and tutorial chapters on their own, then individually attempt to answer the questions and construct working circuits according to the experiment and project guidelines. The instructor never lectures, but instead meets regularly with each individual student to review their progress, answer questions, identify misconceptions, and challenge the student to new depths of understanding through further questioning. Regular meetings between instructor and student should resemble a Socratic⁵ dialogue, where questions serve as scalpels to dissect topics and expose assumptions. The student passes each module only after consistently demonstrating their ability to logically analyze and correctly apply all major concepts in each question or project/experiment. The instructor must be vigilant in probing each student’s understanding to ensure they are truly *reasoning* and not just *memorizing*. This is why “Challenge” points appear throughout, as prompts for students to think deeper about topics and as starting points for instructor queries. Sometimes these challenge points require additional knowledge that hasn’t been covered in the series to answer in full. This is okay, as the major purpose of the Challenges is to stimulate analysis and synthesis on the part of each student.

The instructor must possess enough mastery of the subject matter and awareness of students’ reasoning to generate their own follow-up questions to practically any student response. Even completely correct answers given by the student should be challenged by the instructor for the purpose of having students practice articulating their thoughts and defending their reasoning. Conceptual errors committed by the student should be exposed and corrected not by direct instruction, but rather by reducing the errors to an absurdity⁶ through well-chosen questions and thought experiments posed by the instructor. Becoming proficient at this style of instruction requires time and dedication, but the positive effects on critical thinking for both student and instructor are spectacular.

An inspection of these learning modules reveals certain unique characteristics. One of these is a bias toward thorough explanations in the tutorial chapters. Without a live instructor to explain concepts and applications to students, the text itself must fulfill this role. This philosophy results in lengthier explanations than what you might typically find in a textbook, each step of the reasoning process fully explained, including footnotes addressing common questions and concerns students raise while learning these concepts. Each tutorial seeks to not only explain each major concept in sufficient detail, but also to explain the logic of each concept and how each may be developed

⁴In a traditional teaching environment, students first encounter new information via *lecture* from an expert, and then independently apply that information via *homework*. In an “inverted” course of study, students first encounter new information via *homework*, and then independently apply that information under the scrutiny of an expert. The expert’s role in lecture is to simply *explain*, but the expert’s role in an inverted session is to *challenge*, *critique*, and if necessary *explain* where gaps in understanding still exist.

⁵Socrates is a figure in ancient Greek philosophy famous for his unflinching style of questioning. Although he authored no texts, he appears as a character in Plato’s many writings. The essence of Socratic philosophy is to leave no question unexamined and no point of view unchallenged. While purists may argue a topic such as electric circuits is too narrow for a true Socratic-style dialogue, I would argue that the essential thought processes involved with scientific reasoning on *any* topic are not far removed from the Socratic ideal, and that students of electricity and electronics would do very well to challenge assumptions, pose thought experiments, identify fallacies, and otherwise employ the arsenal of critical thinking skills modeled by Socrates.

⁶This rhetorical technique is known by the Latin phrase *reductio ad absurdum*. The concept is to expose errors by counter-example, since only one solid counter-example is necessary to disprove a universal claim. As an example of this, consider the common misconception among beginning students of electricity that voltage cannot exist without current. One way to apply *reductio ad absurdum* to this statement is to ask how much current passes through a fully-charged battery connected to nothing (i.e. a clear example of voltage existing without current).

from “first principles”. Again, this reflects the goal of developing clear and independent thought in students’ minds, by showing how clear and logical thought was used to forge each concept. Students benefit from witnessing a model of clear thinking in action, and these tutorials strive to be just that.

Another feature of these learning modules is that they do not treat topics in isolation. Rather, important concepts are introduced early in the series, and appear repeatedly as stepping-stones toward other concepts in subsequent modules. This helps to avoid the “compartmentalization” of knowledge, demonstrating the inter-connectedness of concepts and simultaneously reinforcing them. Each module is fairly complete in itself, reserving the beginning of its tutorial to a review of foundational concepts.

To high standards of education,

Tony R. Kuphaldt

Appendix C

Tools used

I am indebted to the developers of many open-source software applications in the creation of these learning modules. The following is a list of these applications with some commentary on each.

You will notice a theme common to many of these applications: a bias toward *code*. Although I am by no means an expert programmer in any computer language, I understand and appreciate the flexibility offered by code-based applications where the user (you) enters commands into a plain ASCII text file, which the software then reads and processes to create the final output. Code-based computer applications are by their very nature *extensible*, while WYSIWYG (What You See Is What You Get) applications are generally limited to whatever user interface the developer makes for you.

The GNU/Linux computer operating system

There is so much to be said about Linus Torvalds' **Linux** and Richard Stallman's **GNU** project. First, to credit just these two individuals is to fail to do justice to the *mob* of passionate volunteers who contributed to make this amazing software a reality. I first learned of **Linux** back in 1996, and have been using this operating system on my personal computers almost exclusively since then. It is *free*, it is completely *configurable*, and it permits the continued use of highly efficient **Unix** applications and scripting languages (e.g. shell scripts, Makefiles, **sed**, **awk**) developed over many decades. **Linux** not only provided me with a powerful computing platform, but its open design served to inspire my life's work of creating open-source educational resources.

Bram Moolenaar's **Vim** text editor

Writing code for any code-based computer application requires a *text editor*, which may be thought of as a word processor strictly limited to outputting plain-ASCII text files. Many good text editors exist, and one's choice of text editor seems to be a deeply personal matter within the programming world. I prefer **Vim** because it operates very similarly to **vi** which is ubiquitous on **Unix/Linux** operating systems, and because it may be entirely operated via keyboard (i.e. no mouse required) which makes it fast to use.

Donald Knuth's \TeX typesetting system

Developed in the late 1970's and early 1980's by computer scientist extraordinaire Donald Knuth to typeset his multi-volume magnum opus *The Art of Computer Programming*, this software allows the production of formatted text for screen-viewing or paper printing, all by writing plain-text code to describe how the formatted text is supposed to appear. \TeX is not just a markup language for documents, but it is also a Turing-complete programming language in and of itself, allowing useful algorithms to be created to control the production of documents. Simply put, *\TeX is a programmer's approach to word processing*. Since \TeX is controlled by code written in a plain-text file, this means anyone may read that plain-text file to see exactly how the document was created. This openness afforded by the code-based nature of \TeX makes it relatively easy to learn how other people have created their own \TeX documents. By contrast, examining a beautiful document created in a conventional WYSIWYG word processor such as Microsoft **Word** suggests nothing to the reader about *how* that document was created, or what the user might do to create something similar. As Mr. Knuth himself once quipped, conventional word processing applications should be called WYSIAYG (What You See Is *All* You Get).

Leslie Lamport's \LaTeX extensions to \TeX

Like all true programming languages, \TeX is inherently extensible. So, years after the release of \TeX to the public, Leslie Lamport decided to create a massive extension allowing easier compilation of book-length documents. The result was \LaTeX , which is the markup language used to create all ModEL module documents. You could say that \TeX is to \LaTeX as **C** is to **C++**. This means it is permissible to use any and all \TeX commands within \LaTeX source code, and it all still works. Some of the features offered by \LaTeX that would be challenging to implement in \TeX include automatic index and table-of-content creation.

Tim Edwards' **Xcircuit** drafting program

This wonderful program is what I use to create all the schematic diagrams and illustrations (but not photographic images or mathematical plots) throughout the ModEL project. It natively outputs PostScript format which is a true vector graphic format (this is why the images do not pixellate when you zoom in for a closer view), and it is so simple to use that I have never had to read the manual! Object libraries are easy to create for **Xcircuit**, being plain-text files using PostScript programming conventions. Over the years I have collected a large set of object libraries useful for drawing electrical and electronic schematics, pictorial diagrams, and other technical illustrations.

Gimp graphic image manipulation program

Essentially an open-source clone of Adobe's **PhotoShop**, I use **Gimp** to resize, crop, and convert file formats for all of the photographic images appearing in the **ModEL** modules. Although **Gimp** does offer its own scripting language (called **Script-Fu**), I have never had occasion to use it. Thus, my utilization of **Gimp** to merely crop, resize, and convert graphic images is akin to using a sword to slice bread.

SPICE circuit simulation program

SPICE is to circuit analysis as **T_EX** is to document creation: it is a form of markup language designed to describe a certain object to be processed in plain-ASCII text. When the plain-text "source file" is compiled by the software, it outputs the final result. More modern circuit analysis tools certainly exist, but I prefer **SPICE** for the following reasons: it is *free*, it is *fast*, it is *reliable*, and it is a fantastic tool for *teaching* students of electricity and electronics how to write simple code. I happen to use rather old versions of **SPICE**, version 2g6 being my "go to" application when I only require text-based output. **NGSPICE** (version 26), which is based on Berkeley **SPICE** version 3f5, is used when I require graphical output for such things as time-domain waveforms and Bode plots. In all **SPICE** example netlists I strive to use coding conventions compatible with all **SPICE** versions.

Andrew D. Hwang's ePiX mathematical visualization programming library

This amazing project is a **C++** library you may link to any **C/C++** code for the purpose of generating PostScript graphic images of mathematical functions. As a completely free and open-source project, it does all the plotting I would otherwise use a Computer Algebra System (CAS) such as **Mathematica** or **Maple** to do. It should be said that **ePiX** is *not* a Computer Algebra System like **Mathematica** or **Maple**, but merely a mathematical *visualization* tool. In other words, it won't determine integrals for you (you'll have to implement that in your own **C/C++** code!), but it can graph the results, and it does so beautifully. What I really admire about **ePiX** is that it is a **C++** programming library, which means it builds on the existing power and toolset available with that programming language. Mr. Hwang could have probably developed his own stand-alone application for mathematical plotting, but by creating a **C++** library to do the same thing he accomplished something much greater.

`gnuplot` mathematical visualization software

Another open-source tool for mathematical visualization is `gnuplot`. Interestingly, this tool is *not* part of Richard Stallman’s GNU project, its name being a coincidence. For this reason the authors prefer “gnu” *not* be capitalized at all to avoid confusion. This is a much “lighter-weight” alternative to a spreadsheet for plotting tabular data, and the fact that it easily outputs directly to an X11 console or a file in a number of different graphical formats (including PostScript) is very helpful. I typically set my `gnuplot` output format to default (X11 on my Linux PC) for quick viewing while I’m developing a visualization, then switch to PostScript file export once the visual is ready to include in the document(s) I’m writing. As with my use of `Gimp` to do rudimentary image editing, my use of `gnuplot` only scratches the surface of its capabilities, but the important points are that it’s *free* and that it *works well*.

Python programming language

Both Python and C++ find extensive use in these modules as instructional aids and exercises, but I’m listing Python here as a *tool* for myself because I use it almost daily as a *calculator*. If you open a Python interpreter console and type `from math import *` you can type mathematical expressions and have it return results just as you would on a hand calculator. Complex-number (i.e. *phasor*) arithmetic is similarly supported if you include the complex-math library (`from cmath import *`). Examples of this are shown in the Programming References chapter (if included) in each module. Of course, being a fully-featured programming language, Python also supports conditionals, loops, and other structures useful for calculation of quantities. Also, running in a console environment where all entries and returned values show as text in a chronologically-ordered list makes it easy to copy-and-paste those calculations to document exactly how they were performed.

Appendix D

Creative Commons License

Creative Commons Attribution 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License (“Public License”). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. **Adapter’s License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

d. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

e. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or

limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

f. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

g. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

h. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.

i. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

j. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

k. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 – Scope.

a. License grant.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

A. reproduce and Share the Licensed Material, in whole or in part; and

B. produce, reproduce, and Share Adapted Material.

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures.

For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.

5. Downstream recipients.

A. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

B. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. Other rights.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.

2. Patent and trademark rights are not licensed under this Public License.

3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. Attribution.

1. If You Share the Licensed Material (including in modified form), You must:

A. retain the following if it is supplied by the Licensor with the Licensed Material:

i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);

ii. a copyright notice;

- iii. a notice that refers to this Public License;
- iv. a notice that refers to the disclaimer of warranties;
- v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.

4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

- a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;
- b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and
- c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors,

whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or
2. upon express reinstatement by the Licensor.

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully

be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the “Licensor.” Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark “Creative Commons” or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at creativecommons.org.

Appendix E

References

“Antenna Deployment Technical Brief”, Hewlett-Packard Development Company, 2006.

Balanis, Constantine A., *Antenna Theory – Analysis and Design*, Third Edition, John Wiley & Sons Incorporated, New Jersey, 2005.

The ARRL Antenna Book, Eleventh Edition, The American Radio Relay League, Inc., Newington, CT, 1968.

The ARRL Handbook For Radio Amateurs, 2001 Edition, ARRL – the national association for Amateur Radio, Newington, CT, 2001.

“CB Radio Service Rules”, FCC Part 95, Subpart D, document U58UT396ZZZ, Federal Communications Commission.

Code of Federal Regulations, Title 47, Volume 1, Chapter 1, Part 15, Subpart C, “Intentional Radiators”, document 47CFR15.247, pages 733-735, revised 2001.

Field Antenna Handbook, U.S. Marine Corps document MCRP 6-22D, 1999.

Hornby, Gregory S.; Lohn, Jason D.; Linden, Derek S., “Computer-Automated Evolution of an X-Band Antenna for NASA’s Space Technology 5 Mission”, preprint for IEEE TEC, 30 November 2007.

Smith, W. W., *The “Radio” Handbook*, Sixth Edition, Radio Ltd., Santa Barbara, CA, 1939.

“Understanding The FCC Regulations for Low-Power, Non-Licensed Transmitters”, OET Bulletin number 63, Office of Engineering and Technology, Federal Communications Commission, October 1993.

Young, Michael F., “Planning a Microwave Radio Link”, YDI Wireless, Falls Church, VA, 2002.

Zyren, Jim and Petrick, Al, “Tutorial on Basic Link Budget Analysis”, Intersil Application Note AN9804.1, 1998.

Appendix F

Version history

This is a list showing all significant additions, corrections, and other edits made to this learning module. Each entry is referenced by calendar date in reverse chronological order (newest version first), which appears on the front cover of every learning module for easy reference. Any contributors to this open-source document are listed here as well.

22 July 2025 – minor edits to the text describing how right-hand circular polarization works on matching antenna pairs.

21-22 June 2025 – added approximate impedance value for 1/2 wave loop antenna to image_4246, and also impedance comment to image_4242. Made some clarifying comments about “RF” early on in the Tutorial, and also edited the text to avoid assumptions that resonance is always the goal for antenna operations. Added comments and a photograph concerning gamma-match networks used in Yagi arrays to increase input impedance. Also created a new section in the Tutorial specifically for standing wave ratio (SWR), and a brand-new section describing voltage and current distributions for a dipole antenna. Also made small edits to section titles within the Tutorial. Also created a new antenna subsection on circularly-polarized antennas.

17 March 2025 – added comments to the Tutorial explaining how dipole antennas have about half the gain as monopoles.

6 March 2025 – added comments to the Tutorial explaining how Yagi antennas achieve directionality, as well as earlier comments on the phenomenon of antenna *gain* prior to the dedicated section on that topic. Also moved the “Antenna reciprocity” section to an earlier place in the Tutorial, as this concept is now mentioned in the “Antenna characteristics” section.

22 October 2024 – added a new Case Tutorial section with example decibel calculations for RF cable losses.

9 October 2024 – added a bullet-point on decibels to the “Challenging concepts” section of the Introduction chapter.

30 September 2024 – divided the Introduction chapter into sections, one with recommendations for

students, one with a listing of challenging concepts, and one with recommendations for instructors. Also added Technical Reference sections on the Poynting vector and Maxwell's equations.

19-20 February 2024 – added a new section to the Technical References chapter on near-field versus far-field effects. Also added a new Conceptual Reasoning question.

24 October 2023 – commented on relative antenna gains for dipole versus ideal monopole.

2 June 2023 – added a Quantitative Reasoning question challenging students to complete a computer program written in C to convert between wavelength and frequency.

28-29 November 2022 – placed questions at the top of the itemized list in the Introduction chapter prompting students to devise experiments related to the tutorial content.

7-9 May 2022 – added a Case Tutorial chapter, with a section showing a home-made 900 MHz antenna.

23 April 2022 – clarified some details on radiation patterns of loop antennas, and moved this section ahead of the Yagi antenna section so that I could discuss folded dipoles as Yagi driven elements. Edited image_4242 and image_4244 for Yagi antennas. Added a subsection on corner reflector antennas. Added more index entries for antenna types.

20 April 2022 – added some content to the Tutorial on log-periodic antennas.

9 March 2022 – added some content to the Introduction chapter, and also made minor edits to some of the antenna.

12 January 2022 – added folded dipole antennas to the “Loop antenna” subsection.

7 January 2022 – changed the name of the “Antenna radiation patterns” section to “antenna characteristics” and then divided it into subsections for each antenna type. Also included comments on radiation resistance and balanced/unbalanced for each antenna type (text within the images).

10-13 October 2021 – added a dimension showing half-wave ($\frac{\lambda}{2}$) length of a Yagi antenna's active element, in image_4237. Also, made minor edits to illustrations.

19 September 2021 – added “unity fraction” unit-conversion example to the “Antenna dimensions” Quantitative Reasoning question.

29 May 2021 – added photograph of a monopole antenna with counterpoise.

13 May 2021 – added graphic images and text about loading coils to the Tutorial, as well as explanations of “trimming” an antenna for resonance.

10 May 2021 – commented out or deleted empty chapters.

4 May 2021 – added comments to the Tutorial about the importance of monopole antenna grounding and counterpoises.

14-15 April 2021 – added a Quantitative Reasoning question on electric field strength versus estimated transmitter power, and also corrected some capitalization errors (“volts” should have been “Volts”). Added more content to some questions, and made minor edits to the Tutorial including “elevation plane” and “azimuth plane” labels to antenna radiation pattern illustrations. Corrected error on loop antenna radiation pattern (I had azimuth and elevation views swapped).

18 November 2020 – significantly edited the Introduction chapter to make it more suitable as a pre-study guide and to provide cues useful to instructors leading “inverted” teaching sessions. Also made some minor edits to the Tutorial.

3 September 2020 – added a Conceptual Reasoning question regarding a passive microwave detector’s orientation versus a Wi-Fi router.

27 July 2020 – document first created.

Index

- Adding quantities to a qualitative problem, 94
- Ampère's Circuital Law, 65
- Ampère's Law, 56
- Ampère, André-Marie, 65
- Annotating diagrams, 93
- Antenna, 67
- Antenna directionality, 24
- Antenna gain, 24, 29, 32, 34, 37
- Antenna impedance, 29
- Antenna impedance, dipole, 22
- Antenna impedance, folded dipole, 23
- Antenna polarization, 24
- Antenna reciprocity, 29
- Antenna, corner reflector, 36
- Antenna, dipole, 17, 31
- Antenna, folded dipole, 32, 35, 36
- Antenna, half-wave, 17, 20
- Antenna, helical, 38
- Antenna, log-periodic dipole, 37
- Antenna, loop, 32
- Antenna, monopole, 17, 20, 30
- Antenna, planar, 39
- Antenna, quarter-wave, 17, 20, 30
- Antenna, turnstile, 38
- Antenna, whip, 17, 20, 30
- Antenna, Yagi, 17, 25, 33
- Bel, 46
- Capacitance, 18, 19
- CB radio, 21, 43
- Characteristic impedance, 43
- Characteristic impedance of free space, 58
- Checking for exceptions, 94
- Checking your work, 94
- Citizen's Band, 21
- Citizen's Band radio, 43
- Code, computer, 103
- Coil, loading, 21
- Common logarithm, 46
- Conservation of Energy, 34, 41, 61
- Constructive interference, 34
- Corner reflector antenna, 36
- Counterpoise, 21
- Cross product, 60
- dB, 40, 46
- dBm, 50
- dBW, 51
- Decibel, 40, 46
- Destructive interference, 34
- Dimensional analysis, 93
- Dipole antenna impedance, 22
- Dipole antenna, half-wave, 17, 20
- Directionality, antenna, 24
- Edwards, Tim, 104
- Electromagnetic wave, 57, 63
- Far field region, 57
- Faraday's Law of Electromagnetic Induction, 56
- Faraday's Law of electromagnetic induction, 65
- Faraday, Michael, 65
- Folded dipole antenna, 32, 35, 36
- Folded dipole antenna impedance, 23
- Fourier transform, 52
- Fundamental frequency, 18, 20
- Gain, 40
- Gain, amplifier, 46
- Gain, antenna, 24, 29, 32, 34, 37
- Gamma match, 35
- Gauss's Law of electric fields, 64
- Gauss's Law of magnetic fields, 64
- Graph values to solve a problem, 94

- Greenleaf, Cynthia, 69
- Ground plane, 21
- Half-wave dipole antenna, 17, 20, 31
- Harmonic frequency, 18, 20
- Heaviside, Oliver, 56, 64
- Helical antenna, 38
- Hertz, Heinrich, 66
- How to teach with these modules, 101
- Hwang, Andrew D., 105
- Identify given data, 93
- Identify relevant principles, 93
- Impedance of free space, 58
- Impedance, antenna, 29
- Impedance, characteristic, 43
- Impedance, surge, 43
- Inductance, 18, 19
- Induction coil, 67
- Interference, constructive, 34
- Interference, destructive, 34
- Intermediate results, 93
- Inverse-Square Law, 58
- Inverted instruction, 101
- Isotropic antenna, 40
- Joule's Law, 32, 53, 60
- Knuth, Donald, 104
- Lamport, Leslie, 104
- Laplace transform, 52
- Lenz's Law, 56
- LHCP, 39
- Limiting cases, 94
- LNA, 42
- Loading coil, 21
- Log-periodic dipole antenna, 37
- Logarithm, common, 46
- Loop antenna, 32
- Low noise amplifier, 42
- Maxwell's electromagnetic equations, 56, 64
- Maxwell, James Clerk, 56, 64
- Metacognition, 74
- Monopole, 64
- Monopole antenna, 20
- Monopole antenna, quarter-wave, 17, 30
- Moolenaar, Bram, 103
- Murphy, Lynn, 69
- Near field region, 57
- Open, 22
- Open-source, 103
- Parasitic effect, 18, 19
- Planar antenna, 39
- Polarization, antenna, 24
- Poynting vector, 57
- Poynting, John H., 60
- Problem-solving: annotate diagrams, 93
- Problem-solving: check for exceptions, 94
- Problem-solving: checking work, 94
- Problem-solving: dimensional analysis, 93
- Problem-solving: graph values, 94
- Problem-solving: identify given data, 93
- Problem-solving: identify relevant principles, 93
- Problem-solving: interpret intermediate results, 93
- Problem-solving: limiting cases, 94
- Problem-solving: qualitative to quantitative, 94
- Problem-solving: quantitative to qualitative, 94
- Problem-solving: reductio ad absurdum, 94
- Problem-solving: simplify the system, 93
- Problem-solving: thought experiment, 93
- Problem-solving: track units of measurement, 93
- Problem-solving: visually represent the system, 93
- Problem-solving: work in reverse, 94
- Qualitatively approaching a quantitative problem, 94
- Quarter-wave antenna, 17, 20, 30
- Radiator, 66, 67
- Radio frequency, 17
- Reading Apprenticeship, 69
- Reciprocity, antenna, 29
- Reductio ad absurdum, 94, 100, 101
- Region, far-field, 57
- Region, near-field, 57
- Resonance, 19, 21
- Resonator, 66, 67

- RF, [17](#)
- RHCP, [39](#)
- Right-hand rule, [60](#)

- Schoenbach, Ruth, [69](#)
- Scientific method, [74](#)
- Short, [23](#)
- Simplifying a system, [93](#)
- Socrates, [100](#)
- Socratic dialogue, [101](#)
- SPICE, [69](#)
- Stallman, Richard, [103](#)
- Standing wave ratio, [43](#)
- Standing waves, [43](#)
- Surge impedance, [43](#)
- SWR, [43](#)

- Tank circuit, [21](#)
- Thought experiment, [93](#)
- Torvalds, Linus, [103](#)
- Transform function, [52](#)
- Transmission line, [22](#)
- Turnstile antenna, [38](#)

- Units of measurement, [93](#)

- Vector cross-product, [60](#)
- Vector, Poynting, [57](#)
- Visualizing a system, [93](#)
- Voltage standing wave ratio, [43](#)
- VSWR, [43](#)

- Wave, electromagnetic, [57](#)
- Whip antenna, [17](#), [20](#), [30](#)
- Work in reverse to solve a problem, [94](#)
- WYSIWYG, [103](#), [104](#)

- Yagi antenna, [17](#), [25](#), [33](#)
- Yagi, Hidetsugu, [25](#), [33](#)