Math and Physics for the 802.11 Wireless LAN Engineer

A Discussion of What Every LAN Engineer Should Know About 802.11

by Joseph Bardwell

Math and Physics for the 802.11 Wireless LAN Engineer" is a discussion of physics and electromagnetic wave theory as applied to 802.11 wireless networking. Mr. Joseph Bardwell has written a readable paper that provides an explanation of how Maxwell's wave equations, Fresnel Zone calculations, and many other complicated engineering topics can be readily understood, and how they come into play in the realm of wireless network design and implementation. While there is an allusion to Calculus, the paper is written so that anyone with a high school algebra background can easily follow the math. You'll learn about how electromagnetic waves are affected by the environment and you'll be introduced to some of the more esoteric quantum mechanical characteristics of particle/wave duality. Ever wonder how a radio signal can go around the corner of a building in the absence of any reflective surfaces to bounce from? You'll find out.

October 2003

WildPackets

Table of Contents

Math and Physics for the 802.11 Wireless LAN Engineer

About the Author

Joseph Bardwell has been active in the computer industry since the 1970's and is a widely recognized lecturer, technical consultant, and co-author of the recent book "Troubleshooting Campus Networks" (Wiley, 2002). He has been a contributor to the development of the WildPackets AiroPeek NX^{tm} wireless LAN analyzer and the original Network General Sniffer^{tm} protocol analyzer. He was named as one of the "Service 25" a list of key, influential people in the network services industry for his pioneering work in developing technical certification for network analysis. Over the past several years hundreds of people across the country have heard him speak in the "ZEN and the Art of Network Maintenance" seminar series that he developed. As Mr. Bardwell says in his seminars "When you know what the magician knows, it's not magic anymore."

Mr. Bardwell is the Chief Scientist and President of Connect802 Corporation (www.Connect802.com), providing turnkey Wi-Fi networking solutions and predictive RF modeling and simulation for 802.11 network design. He is a certified RF designer. Mr. Bardwell may be contacted via email at: joe@Connect802.com.

Section 1: Introduction

"I'm going to let my chauffeur answer that…"

A prominent professor was asked to give a lecture on a highly complex technical topic to audiences at a number of different universities around the country. He was uncomfortable flying and so a limousine and a chauffeur were hired to drive him from location to location. After several weeks on the road the chauffeur and the professor were on their way to the next city. The chauffeur said "Professor, you are a very smart man and you earn a professor's salary and I'm a chauffeur and I earn a chauffeur's salary. I've been listening to you give the same lecture, night after night, for weeks now and, frankly, I could give the same speech." The professor replied "At the university where I'm speaking tonight nobody has ever met me. We're about the same size and build; stop here at this gas station and we'll exchange clothes. You'll wear my suit and give the lecture tonight and I'll sit quietly in the back row wearing your chauffeur's uniform."

The professor sat in the back row and the chauffeur gave the lecture. The speech was delivered perfectly and culminated in a rousing round of applause. One fellow in the audience raised his hand and asked "Professor, does the locus of the covariant tensor have a non-commutative divergence in the field of the transfinite singularity?" The answer he got was "Sir, I can't believe they would allow a person with your obvious lack of education into this university. That question is so basic, so simple, why… I'm going to let my chauffeur in the back row answer it!"

Are You the Professor, or the Chauffeur?

A good deal of properly designing, implementing, and troubleshooting an 802.11 wireless LAN is based on the physics of electromagnetic field propagation. There are some folks who will tell you how much they know about 802.11 wireless networking and how they can help you design, implement, or troubleshoot your system. Some of these people know what they're talking about; others are, perhaps, just talking. It doesn't take long for a LAN engineer, experienced with Ethernet and TCP/IP to learn the basics of 802.11 networking and come up with all the right words at just the right times. Saying the right words is easy. Understanding the details of radio signal propagation and wireless network security is challenging. Not everyone who can "talk the talk" is able to "walk the walk". "You need a site survey" says one person. "802.11i and WPA will secure your network" says another. "Use a directional antenna with lots of gain" chimes in someone else. If you know the fundamentals of the underlying physics then you're on your way to knowing who is the professor, and who is the chauffeur.

Since you're reading this, it might be assumed that somewhere in the labyrinth of your mind there was a haunting recognition that the title "Math and Physics for the 802.11 wireless LAN engineer" struck a resonant chord with your own sense of curiosity. If you've thought about how an 802.11 wireless transmitter manages to convey a signal through space to a receiving station and you realized that there was more to it than "…channel 1, 6, and 11 are non-overlapping" then your choice of reading material is correct!

Purpose and Perspective

Unless you're armed with accurate technical information, you aren't equipped to effectively consider, design, implement, or troubleshoot any new technology. This has been historically true in the computer industry and it's true today as the new 802.11 wireless LAN technologies become pervasive in the marketplace.

The computer industry has been a consistent focal point for much of the world since the time (in the 1950's) when a UNIVAC computer correctly predicted the outcome of the U.S. presidential race. Corporate management was able to see the benefits of early computerization efforts in the workplace but, for the most part, they were ignorant of the underlying engineering that made the machines work. Ignorance often creates fear and myth and, early on, IBM met the concerns of corporate decision makers by installing a big, red lever switch in the computer room. It was isolated in a glass-front red box with a "break glass" hammer hanging next to it. You see, there were concerns that the new computer technology could, in some mysterious way, get out of control and do something bad, perhaps taking over with a mechanical mind of its own. IBM said "Don't worry, if there's a problem all you have to do is break the glass, pull the big red lever, and turn the whole darn thing off!"

In 1981, when the IBM Personal Computer became the touchstone for the future of desktop computing around the world, the tradition was continued (albeit without much fanfare) and the IBM PC sported a big, flat, red lever on the side with which the darn thing could be turned on and off.

The computer industry evolved to become an economic and cultural force that brought changes to society rivaled only, perhaps, by the Industrial Revolution, or other pivotal events in the history of the human race. With the creation of the World Wide Web, and the pervasive penetration of computer networking into the business, educational, and government sectors, the Internet has become one of the most significant manifestations of technology in the world today. Nonetheless, the proverbial "digital divide" extends to society as a whole along with the business managers and decision makers who must depend on highly technical engineers for information about how to best take part in the technology explosion. In some sense this is a case of "the blind leading the blind."

Computer systems engineers, software developers, technical support representatives, and all manner of specialized titles are charged with a daunting responsibility. Each person, as they strive to meet the goals of their job function, must necessarily have a strong focus on the equipment, software, and methods that are currently deployed in their environment. As a result, when new technology is introduced the general population of technical experts is, for some period of time, ignorant of the features, benefits, and potential problems inherent in the new, sophisticated offerings. When a new technology is a natural extension of an existing technology, the period of ignorance is very short.

Consider the late 1980's. Network engineers who were experts in the design, implementation, and troubleshooting of coaxial Ethernet were faced with the learning curve associated with the nuances of twisted-pair Ethernet and, shortly thereafter, switched Ethernet. Rolled into this were considerations relating to full-duplex transmission and fiber optic networking. Ultimately, however, Ethernet was Ethernet and the familiar aspects of Version 2 and 802.3 implementation, coupled with the omnipresence of TCP/IP, allowed the transition to occur with a minimum of confusion. There were, however, many years (in the early 1990's) when the concepts of "Layer 2 Switching" versus "Layer 3 Switching" and even "Layer 4 Switching" were often mind boggling due to the lack of ratified and implemented standards, vendor impatience and desire to get new products to market, and the need for marketing and advertising folks to come up with succinct sound bites to describe complex engineering in terms that sales people could use.

The introduction of low-cost, readily available, commercial wireless LAN equipment at the turn of the new millennium has, once again, formed a point of demarcation in the growth and evolution of the computer industry. This time, however, the new technology is not a natural evolution of existing technology. A statement like that needs some clarification and the clarification makes potential pitfalls even more subtle. It should be obvious that microwave transmission has been around for quite some time. Electronics engineers and designers of radio transmission and reception equipment have been plying their trade for over 100 years in one form or another. There's nothing inherently new about transmitting bits in the 2.4 GHz or 5.8 GHz Industrial, Scientific, and Medical (ISM) frequency band as is done in an 802.11 wireless LAN. There's nothing inherently new about implementing TCP/IP in a wireless environment, either. There are surely new product offerings for Virtual Private Network (VPN) implementation, and security is a hot topic, but at the end of the day, an IP client is talking to an IP server with good old fashioned DHCP, DNS, ARP, ICMP, and other familiar protocols.

The idea that 802.11 wireless LANs introduce technology that is not a natural evolution from previous technologies comes from the fact that the marriage of radio engineering and LAN engineering brings together two disciplines that have had only limited contact in the past. The most experienced RF (radio frequency) engineers have come up through the ranks, from their academic backgrounds to their present roles in research and industry, focused on electromagnetic wave propagation in one form or another. LAN designers, engineers, and troubleshooters can rate their expertise, in part, on the basis of their skills related to the TCP/IP protocol family. The RF expert is, in almost every case, not the LAN expert and most LAN experts are not, with very few exceptions, knowledgeable regarding the intricacies of electromagnetism, antenna design, or atmospheric signal propagation characteristics. For this reason there is a potential chasm facing the person responsible for implementing a wireless LAN. They have to know not only how to design the LAN elements of the network, but they're also faced with RF engineering issues that can verge on having incomprehensible foundations.

Apprehensive Attitudes Resulting from Lack of Knowledge

When the LAN engineer, designer, troubleshooter, or management decision maker is faced with RF engineering questions about which they know only what some vendor's marketing literature tells them, they find themselves to be ignorant of many fundamental facts. Ignorance, then, yields fear and, in the 802.11 wireless LAN marketplace, the fears run wild. Some examples of the apprehensive "self talk" that might go through a person's head as they contemplate an 802.11 implementation might sound like the following.

"An intruder might be sitting in the parking lot listening to, hacking in to, or otherwise doing something naughty with my wireless network."
"My 2.4 GHz cordless phones, microwave ovens in the kitchen, and all manner of mysterious environmental influences may render my investment in 802.11 infrastructure useless or severely devalued."
"I don't know exactly how the RF signal is bouncing around in my building, so where should I install 802.11 wireless access points to most effectively provide high-speed, reliable coverage?"
"I hear about 802.11b, 802.11a, and 802.11g and a whole bunch of other standards, too. I can't make sense out of what the features and benefits are, exactly, for each option, so how can I properly design and implement a wireless LAN?"
"I've been told that consulting companies can perform site surveys to ascertain the specific RF characteristics of my building but, when I found out what a site survey would cost, it was prohibitively expensive!"
"Vendors provide detailed specifications for their equipment and test tools, but it seems like someone would have to be a physicist to fully appreciate what all those specs mean, and how they might impact my implementation."
"There are so many products, tools, and support services into which I could throw money, but I'm

not really sure of exactly how all this stuff works and my best efforts to design a suitable wireless network contains a high degree of guesswork."

"Budgets are tight and I need to know, in advance, what I'm going to have to spend to implement an 802.11 network but nobody seems to be able to give me a straight answer… it's like all of these other technical people are making it up as they go along."

"It seems like there are only a handful of people who really understand the marriage of RF engineering and LAN technology. How did they figure it out?"

"Knowledge is power" is a profound statement. Knowledge of the RF side of wireless LAN technology can help reconcile apprehension to fact and confusion to reality. Today's LAN engineers, designers, and managers must be empowered to implement 802.11 wireless LANs. In the following pages you're going to read about the physical properties of electromagnetic wave propagation as they may be applied to the realm of 802.11 wireless LAN communication. You'll be given the basis for signal strength evaluation, antenna selection and placement, and site survey assessment. The core knowledge that you need to assimilate, while it may be touched on in one or another discussion, goes beyond what any single text could offer. Electromagnetism is a science that begins as one of the four fundamental forces in nature (along side gravity, and the strong and weak nuclear force). Physicists have been studying electromagnetic effects for over 100 years and, although they have developed wonderful theories of Quantum Electrodynamics (QED) and have tried to unify the four fundamental forces, they don't have all the answers yet.

What You'll Learn in this Paper

As you begin reading you'll probably find that some topics are significantly more familiar than others. Hopefully the gaps between what you already know and the mysteries of 802.11 RF engineering can be bridged to some degree, at least in concept. It is hoped that your greater comprehension of how signals are transmitted, propagate through space, and are received will serve as fuel for the furnace of intuition that you ignite when you're in the field, trying to solve a real-world problem or design an 802.11 wireless LAN.

An effort has been made to present the ideas to be discussed in an orderly manner, but you must realize that some of these concepts are consistent with a graduate level course in physics and you may wonder why certain topics have been emphasized. There are many, many books that discuss the laws and principles involved in the transmission of electromagnetic energy and a web-search will quickly provide additional references. It has been said, "Knowledge is power" but, as you'll discover in your reading "Electromotive force multiplied by current flow is power," too.

A Note to the Reader Familiar with the Subject

While many readers will be new to electromagnetic theory, some of you may have a solid background in the topics covered and are interested only in how these are applicable to the 802.11 wireless LAN environment. If this is you, please remember who the primary audience is for this discussion. Many, many topics have, of necessity, been abbreviated and, in some cases, presented in such an oversimplified manner as to make a physicist groan.

The discussion of Maxwell's equations introduces the concept of curl, but glosses over div and grad, without touching on the Laplacian operator. Vector quantities have been used without mention of the dot or cross products. Some allusions to quantum mechanics have been introduced with no mention of the two-slit experiment, the Hamiltonian operator, or matrix calculations in any form. The discussion introduces the retarded wave but neglects to expand on the relativistic effects in general because these effects are unimportant for understanding 802.11 wireless LAN communication. As far as the math is concerned, there are references to derivatives, partial derivatives, and integrals but only with enough clarity to get the salient points across. The whole concept of limits has been left unspoken, and, in like manner, no mention is made of using an algebraic sum inside an integral.

In a 100-page document there's no way to convey the scope and depth of a graduate-level course in physics! When, as a reader with a strong background in calculus or electromagnetism, you stub your mental toe on some particular description, you're encouraged to take a step back and ask yourself "Is this a sufficiently reasonable explanation for the target audience?" It's hoped that some readers will be motivated to pursue further study of the topics that they find intriguing and that they will then gain a broader understanding of the details.

Section 2: Electricity and Electromagnetic Fields

The physical properties of an antenna cause an electromagnetic field to propagate outwards into the surrounding space. To lay a foundation for discussing the details of antenna operation and RF assessment, it's necessary to begin with some details regarding electromagnetic fields. The fields generated by an 802.11 antenna (and, for that matter, all antennae) are the result of energy being conveyed into the antenna through the movement of electrons in the conducting metal parts of the antenna. The electrical force produces an electromagnetic field that is propagated outwards into space. The receiving antenna is influenced by this expanding electromagnetic field and an electrical force (albeit with some degree of loss or distortion) is induced into the electrons in its metal parts.

Electrical Force

Atoms have one or more electrons moving around a central nucleus containing protons and (with the exception of hydrogen) neutrons. We'll leave it up to the quantum physicists to explain the inner workings of the atom, but it's sufficient to know that an electron carries a negative charge and the proton carries a positive charge. Since like charges repel, it should be evident that if two electrons were to approach each other the repulsive force between them would get larger as they got closer together. The same would be true for the attractive force between an electron and a proton: it would increase as the two got closer. The result would be, conceptually speaking, that two electrons would want to move away from each other (repulsive force) more vigorously if they were closer together than if they were farther apart. Some amount of force is, therefore, exerted when two electrons repel each other. This fact gives rise to the need for some engineering definitions of force and work.

The "work" to move an object is given by the simple equation "Work = Force times Distance." Force may be measured in a unit called a *Newton*, which is the metric unit for weight. Remember that kilograms are actually units that measure mass. If an object's mass (in kilograms) is multiplied by the gravitational force of 9.8 m/s2 (meters per second per second) you get the weight in Newtons. For example, 125 pounds is roughly equal to 550 Newtons (1 lb. = .45 kg, then multiply by 9.8). If a particular force (Newtons) is moved through a certain distance (meters) the resulting work unit is called a *Joule* (pronounced "jewel"). One Joule is equal to a force of 1 Newton pushing an object through a distance of 1 meter. In essence one Joule is equivalent to one "Newton meter."

When thinking about the forces related to electrons and their accompanying charges it becomes unreasonable (except for the quantum physicist) to think solely about one single electron. A very large number of electrons are used as a fundamental unit. When 6.24×10^{18} electrons are considered as a single group it's referred to as 1 *Coulomb*. A Coulomb is simply 6,240,000,000,000,000,000 (six million two hundred forty thousand trillion) electrons.

If one Coulomb of electrons has one Joule of work-producing capacity, we say that one *Volt* of electromotive force (EMF) is present. When we say that one volt is present we're referring to a relative measurement. That is, the electrical push, or pressure, that is present in the Coulomb of electrons is pressure relative to something else. Think back to the idea of only two repelling electrons. The repulsive force doesn't exist for one electron alone. The repulsive pressure only exists when a second electron enters the picture. In everyday electrical engineering (terrestrial, as opposed to spacecraft) this standard reference point is the electrical potential of the planet earth itself. This is referred to as the *ground potential*.

In a standard (dry cell) battery (like a "D" cell or 9-volt battery) the chemistry of the materials that are manufactured into the battery result in a net negative charge (extra electrons) being in one part of the cell, and a net positive charge (a lack of electrons) being in another part of the cell. This is an oversimplification, but it's sufficient at this point in the explanation. The system would return to equilibrium (with the electrons being distributed in proportion to the protons; essentially uniformly), except the electrons are blocked from doing so. When a wire is connected from the negative to the positive terminals of the battery, the electrons are provided with a path through which they can return to their state of lower energy. The electrical potential, or "pressure" is caused by the electrons attempting to reestablish their state of chemical equilibrium. A 9-volt battery has nine volts of electromotive differential force between the positive and negative terminals. Again, voltage is a relative measure of electrical potential. The 9 volts in the battery are there because of the electrical difference between the positive and negative sides and not as the result of some inherent characteristic of the material or the electrons themselves.

In an 802.11 radio the circuitry creates an electrical potential in the radiating element of the antenna causing the electrons in the element to move. The moving electrons create an electromagnetic field that propagates away from the antenna. The relationships between the electrical and magnetic characteristics of this field are described in the study of electromagnetism.

Ohm's Law

The electromotive force pushing on some electrons makes them want to move. This repulsive force is conveyed from electron to electron through a conducting medium (like a wire or an antenna) when an energy differential (voltage) is created between two portions of an electrical system. The movement will be resisted by the medium (the material through which they are moving). Resistance to the flow of electrons is measured in a unit called an *Ohm* (represented by the Greek letter Omega, Ω). It's named for the German physicist George Simon Ohm who was born in 1787 and died in 1854. When one volt can move one Coulomb of electrons past a point in one second, the resistance is said to be equal to one Ohm. One Coulomb of electrons moving past a point in one second is called one *Ampere* (or Amp, represented by a capital letter "A"). It's named for Andre Ampere (1775-1836), a French physicist who laid the foundations for the science of electrodynamics by demonstrating, and subsequently studying, the fact that electric currents produce magnetic fields.

It's worth noting that electrons have mass (a very small mass, but they're not "massless" in the way a photon is considered to be a massless particle). When a mass is set in motion it will stay in motion unless acted on by some other force, as stipulated in Newton's laws of motion. Hence, moving electrons manifest the characteristic of momentum. Once moving they must be acted on, and yield their energy of motion in response to, some other force.

Let's review what's been covered so far:

- 1 coulomb = 6.24 X 10^{18} electrons
- 1 volt = 1 coulomb able to exert 1 joule of force
- 1 volt moves 1 coulomb through a resistance of 1 Ohm in 1 second
- When 1 coulomb moves past a point in 1 second it's called 1 Amp of current which is to say: 1 amp = 1 coulomb/second
- Moving electrons have momentum

There is a mathematical relationship between volts, amps, and Ohms that's expressed by *Ohm's Law*. It says: I = E/R where "I" = the current flow measured in Amps "E" is the voltage measured in volts, and "R" is the resistance measured in Ohms.

Resistance and Reactance

Ohm's Law is sufficiently easy to understand and the concepts are consistent within the scope of our discussion. It is not, however, the complete story. It's true that the physical makeup of a conductor (like a wire, an Ethernet cable, or an antenna cable) resists the flow of electrons to some degree, and this "degree of resistance" is measured in Ohms. Ohm's Law, as described thus far, assumes that the flow of current is steady (direct current, DC) and not changing (alternating current, AC) but it also applies to AC circuits.

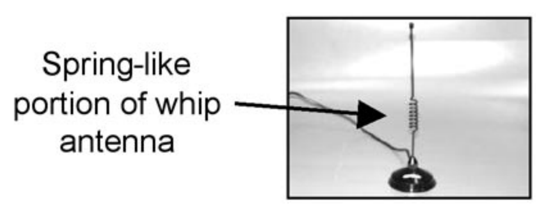
There are two important electrical characteristics that are manifested when current flow varies and these are *capacitance* and *inductance*. The resistance to current flow introduced by these is called *reactance*, and the degree to which current flow is resisted is dependent on the frequency of the alternating current.

Resistance, then, is the opposition to current flow when steady electric currents flow along conductors in one direction. Resistance to alternating current is more complicated and manifests itself in the form of *reactance*. Reactance may be of two types: inductive and capacitive. Inductive reactance is associated with the varying magnetic field that surrounds a conductor whereas capacitive reactance is associated with the changing field between conducting surfaces separated from each other by an insulating medium. While resistance and reactance use the Ohm as their unit of measure, the actual equations relating to reactance involve complex numbers (i.e. the square-root of negative one). The real part of the current becomes a sine function when the real part of the voltage is taken as a cosine function. The consequence that can be revealed in the equations for reactance is:

- Capacitive reactance decreases as frequency increases
- Inductive reactance increases as frequency increases

A resistor opposes alternating current equally as the frequency changes. The degree to which capacitors and inductors oppose alternating current varies with frequency. While the characteristics of capacitance and inductance are present in wires and antennae, circuit board components (discrete components) can be built with pre-specified capacitive or inductive values. Circuit designers use discrete capacitors to block direct current and low frequencies and discrete inductors to pass low frequencies with little or no opposition while blocking higher frequencies.

One practical application of reactance measurements is found in the design of many small "whip" antennae. When you see the spring-like construction at the base of a whip antenna, you're looking at an inductive element designed to offset a capacitive effect introduced by the antenna.



We've taken this side-trip into the realm of capacitive and inductive reactance to provide a perspective on another aspect of electrical engineering. For the discussion of RF signal propagation in the 802.11 environment, a general, basic understanding of Ohm's Law is sufficient.

Power Measurement

There is a special relationship between voltage and current related to the work-producing capacity of a system. From this perspective it can be shown that a large number of electrons (high amperage) with a low electrical pressure (low voltage) can do the same amount of work as a smaller number of electrons (lower current) working across a higher difference in energy potential (higher voltage). The work that can be done by some electrons is directly related to how many of them there are and the voltage level. A circuit with 6 volts that moved 3 amps across a resistance would be doing the same amount of work as one in which 3 volts moved 6 amps. The resistance in the second circuit would, of course, have to be less to allow the lower voltage to push more electrons. In both cases, however, the work done would be the same. The relationship between volts and amps is called electrical *Power* and it's measured in a unit called the *Watt*. One watt is simply 1-volt pushing 1-amp or P=IE (watts of power = current flow in amps multiplied by electrical pressure in volts). A given amount of power can be delivered at a lower voltage with more electrons (low volts, high amps) or at a higher voltage with fewer electrons (high volts, low amps).

A Watt is a unit that measures the rate at which work is done. For example, consider the difference between moving an object using a force of 1 Newton through a medium with very little resistance compared to moving the same object, with the same 1 Newton of force, through a medium with great resistance. In both cases the same mass has been moved with the same force through the same distance. In the second case it took longer (because the medium resisted the movement). The rate at which the work was done was slower in the case of a resistive medium. This is a power measurement. Work done per unit of time (the rate of doing work) is what's being measured. Power (in Watts) is equal to work done (Joules) divided by time. 1 Watt = 1 Joule performed in 1 second. A Watt, therefore, is actually a "Joule per second." A 125-pound person running up the stairs to a height of 3 meters in 2 seconds has used 825 watts of power (550N X 3m) / 2 seconds).

Now, a typical 802.11 access point may be rated at 100 mW (milliwatts) with the output power being equal, therefore, to 0.100 watts. This is a measure of what's called "effective radiated power" or "ERP." ERP is a magical way to "hand wave" over a number of complicated characteristics of signal transmission. The need to hand wave arises from the fact that actual power can only be calculated based on the electrical charges present in the radiating element and in the field pattern.

Watts, Milliwatts, Decibels, and dBm Units of Measurement

A milliwatt (mW) is 1 one-thousandth of a watt or 0.001 watts; that's easily understood. We can't leave the topic of measurement, however, without introducing the "dBm" unit (pronounced "D, B, milliwatts" or simply "D, B, M"). A vendor might specify that their PCMCIA 802.11 adapter has a receive sensitivity (ability to extract bits from the signal) of "-84dBm at 11 Mbps." The dBm unit is a logarithmic unit of power measurement as opposed to mW, which is a linear measurement. Both units are measurements of exactly the same thing, the power level that, ultimately, is some number of milliwatts. It's only a matter of whether the milliwatts are being measured using the linear mW scale or the logarithmic dBm scale.

The reason that a logarithmic scale is used to measure power is that the mW value gets very small very quickly. For example, in a typical office building it might be found that the signal power 150-feet away from an access point is 0.000000000316 mW. That's too many zeros to use in casual conversation (or to easily manipulate in field calculations!). Using the logarithmic dBm, scale this same signal power is represented as - 95 dBm.

The dBm scale is based on the use of decibels and so, to understand dBm, it's first necessary to discuss decibels. Decibels and dBm are based on a common arithmetic basis but they are different types of quantities. Named after Alexander Graham Bell, the actual value is called a *bel*, with a decibel being 1/10th of a bel. The bel value is simply the ratio (a simple fraction) comparing two numbers. The original bel measured the power coming out the end of a telephone circuit as a ratio to the power going in. The resulting fraction was represented as a base-10 logarithm to avoid having all those zeroes to the right of the decimal point. A decibel is the same thing except the logarithmic conversion of the ratio is divided by 10 (again, to make the number smaller and more easily manageable).

It's very important to recognize that a decibel is a dimensionless unit. That's because a decibel is, ultimately, just a fraction. There's no arbitrary implication as to what units are being used for the numerator and denominator. In this regard it's common for decibels to be used as a way to indicate increases in a quantity (fraction greater than 1) or a decrease (fraction less than 1).

Because decibels are dimensionless it becomes important to always specify (either explicitly or implicitly) what units are being used for the numerator and denominator of the ratio that's being expressed. When the denominator is 1 milliwatt then the quantity becomes the dBm. On the other hand, sometimes decibels are used to represent a dimensionless ratio but a letter is appended to indicate what the ratio is comparing. An example of this is the "dBi" measurement, a reference to antenna gain, where the denominator is the power level, in milliwatts that would be observed from a theoretical isotropic radiator (one that radiates in a perfect sphere) and an actual antenna (that always radiates with a field pattern that is something other than a perfect sphere). So, although dBm and dBi both use "dB" and a letter, the former is a scalar quantity measuring power and the latter is a dimensionless ratio comparing measured power to theoretical power. You'll encounter many "dB" values in the realm of RF engineering and 802.11 networking, so be on guard as to whether you're seeing a metric or a ratio.

Milliwatts are a linear scale of measurement. That is, 50 mW is half of 100 mW and the power measurement varies in a straight-line fashion. The dBm unit is logarithmic and the power measurement varies in a logarithmic curve. Using dBm units you'll find that a variance of 3 dBm either doubles or halves the power (depending on whether you're adding or subtracting). A power measurement of 20 dBm is two times greater than a power measurement of 17 dBm, and half as strong as 23 dBm. You should remember this fact, that +3 dBm doubles the power and -3 dBm halves the power.

One math fact to keep in mind is that when the logarithm is taken for a fractional quantity less than 1 (and greater than 0) it's a negative logarithm. This is analogous to (although not exactly the same as) seeing a number represented in scientific notation (for example 3.456×10^{-9}, which is the number 0.000000003456). The two conversion equations to allow conversion between mW and dBm are shown below.

$$\text{dBm} = 10 \text{ Log}_{10} (\text{mW})$$

$$\text{mW} = 10^{(\text{dBm}/10)}$$

Any calculator (like the Window's calculator in scientific mode) can by used to perform conversions. For example, consider a 100 mW access point. If you enter 100 into your calculator and press the LOG button you'll get the base-10 logarithm for 100 (which is 2). Now multiply by 10 and you'll find that 100 mW = 20 dBm. Going the other way, if you start with 20 dBm you can convert back to mW by using the X^Y function on your calculator (first value, x, raised to the second value, y, power). Enter 10 into your calculator. Press the X^Y key and enter the dBm value divided by 10, which is easy enough since you only have to mentally move the decimal point one place to the left. If the dBm value is negative use the +/- key to change the sign of the number. Try this out by converting -95 dBm into mW and your answer should be 0.000000000316 mW.

You can also remember that 20 dBm = 100 mW and that +3 dB doubles the mW and -3 dB halves it. Consequently you know that 17 dBm = 50 mW, 14 dBm = 25 mW, and so forth. More importantly, you know that if you measure -85 dBm in one location during a site survey and you measure -91 dBm in another location the second location has only 25% of the signal strength that the first one has. This is because when you go from -85 dBm to -88 dBm the power goes down by half. Going from -88 dBm to -91 dBm cuts the power in half again, resulting in 25% of the original measurement.

Magnetic Fields

Andre Ampere's work in the study of electric currents and magnetism laid the groundwork for today's science of electrodynamics. In 1820 he announced the discovery that the magnetic needle of a compass moved in response to a nearby electric current flow. He demonstrated that the magnetic field moves away from and around a conducting wire in direct relationship to the direction of current flow through the wire. Ampere's theorem stated the relationship between the strength of a magnetic field and the strength of the electric current producing it.

Magnetic fields develop perpendicularly around an electric current carrying wire (Figure 2.1, below). The perpendicular aspect refers to the fact that the circular, rotating magnetic field is expanding (developing) in a direction that is 900 to the direction that the current is flowing in the wire. They expand as the electric current (the electrons moving in the radiating element) reach their maximum speed. The magnetic characteristic of the field causes electrons to move in a receiving element resulting in the receiver developing an electric current.

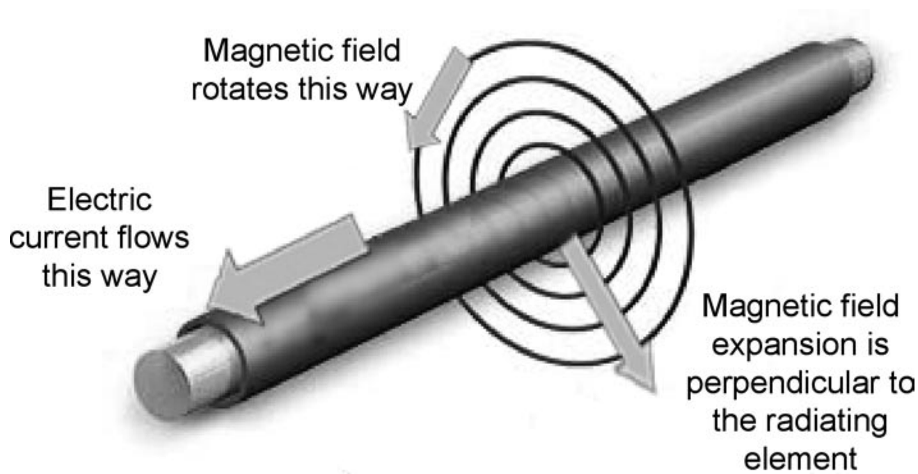


Figure 2.1 The Magnetic Field Surrounding a Current Carrying Conductor

Of course, the field that radiates outward is, itself, a moving electromagnetic field. There is, at each point in this field, an electric component, always at 900 to the magnetic component. As the magnetic field expands it has a rotational characteristic. The direction of rotation can be determined using, what engineers have termed, the “Right Hand Rule.” If you imagine holding the conductor in your right hand, you’ll see that your fingers curl around the conductor in the direction that the field rotates when your thumb points in the direction the current is traveling. In the figure above you would grasp the wire from the top, with your thumb pointing to the left. Your fingers would curl around in the direction shown by the arrow, pointing in the direction that the magnetic field is rotating.

As the magnetic field expands the power radiated into space from the conductor varies inversely with the square of the distance. This is called the *Inverse Square Law*. According to the Inverse Square Law, if you were to measure the power of an antenna at a distance of, say, 10 feet and you found it to be 80 mW then at a distance of 20 feet (distance increased by a factor of 2) you would find the power to be 20 mW (power reduced by a factor of 4). When you’re twice as far away the power is reduced to 1/4 of its previous value. If you were 3 times further away the power would be reduced to 1/9 of its previous value. Power varies inversely as the square of the distance.

The “Inverse Square Law” is fine and good, but it gives rise to a strange conclusion unless it’s properly understood. It’s important to gain a working knowledge of the physical properties that make a magnetic field propagate outward from an 802.11 transmitter and induce current in a receiving antenna. When wireless LAN designers, engineers, or occasionally sales people try to draw valid conclusions about how equipment will interoperate and communicate in a particular environment they are faced with a major stumbling block unless they have the right understanding of how RF signals really behave. An engineer who only has partial knowledge of the physics underlying 802.11 communication can end up very confused. Even the Inverse Square Law, simple enough at first consideration, has a paradox hiding within it. The “Inverse Square Law” could be likened to the famous Greek math puzzles called “Zeno’s Paradoxes.”

Zeno’s Paradoxes

Zeno of Elea was a Greek philosopher who lived around 500 BCE. He proposed four paradoxes that remained unsolved by mathematicians until Cantor’s theory of infinite sets was put forth in the second half of the 19th century. Zeno’s paradoxes describe situations not unlike the movement of force in an electromagnetic field, which relate discrete calculations on individual points to continuous assessment of all of the points together. The distance from an antenna to a measuring point is finite; it’s some precise number of meters. There are, mathematically speaking, an infinite number of points embodied in that distance. Since the Inverse Square Law relates, ultimately, to point-particles in an electromagnetic field, Zeno’s paradoxes are apropos.

Zeno’s first paradox asks for mathematical proof that something can get from Point A to Point B. It’s a paradox because we all know that you can walk from your house to your car, or that a signal can go from an access point to a client computer in a wireless LAN. Nonetheless, the mathematics underlying the proof is not immediately obvious and, in fact, it took hundreds of years before anybody had a clue how to prove it.

To understand Zeno’s first paradox, think about a runner who wants to run 100 meters, in a finite time. To reach a 100-meter mark the runner must first reach the 50-meter mark and, to reach that, the runner must first get to the 25-meter mark. Of course, to get to the 25-meter mark the runner must first get to the 12.5-meter mark. The requirement that the runner must first reach the halfway point before they can reach the next milestone continues infinitely. The runner, therefore, can never start the race since, to reach even the most infinitesimal distance from the starting point they would first have to reach the point halfway there. The runner would have to reach an infinite number of 'midpoints' in a finite time. Infinite points in finite time are impossible, so the runner can never arrive at the finish line. In general, anyone (or any electromagnetic field) wanting to travel from Point A to Point B must meet these requirements, and so motion becomes impossible, and what we perceive as motion is merely a strange illusion.

Zeno's second paradox also suggested a race, but this time between Achilles (a fast athlete) and a turtle. Achilles can run 5 meters per second but the turtle can only run 1 meter per second. The race will be over a 100 meter track. To be fair, Achilles gives the turtle a 5-meter head start. At the sound of the starting gun the turtle is 10 meters ahead of Achilles. After 1-second, Achilles has reached the spot where the turtle started. The turtle, during that one second, has run 1 meter further. Achilles continues running and reaches the 6-meter mark (where the turtle was when Achilles was at the 5-meter mark). Now, however, the turtle has moved away from the 6-meter mark and is, still, ahead of Achilles. This continues and whenever Achilles reaches the spot where the turtle has just been (a fraction of a second earlier) the turtle has again covered a little bit of distance beyond that spot, and is still winning the race. The paradox evolves since it would appear that no matter how hard Achilles tries he can never reach the turtle and the turtle ultimately wins the race.

Bardwell's ERP Paradox

Now, consider "Bardwell's ERP Paradox" (a thought experiment created by the author), that sounds strikingly similar to the problem posed in Zeno's Paradoxes. This, of course, is not to be confused with the famous EPR paradox from 1935, when Albert Einstein and two colleagues, Boris Podolsky and Nathan Rosen (EPR) developed a thought experiment to demonstrate what they felt was a lack of completeness in quantum mechanics.

Bardwell's ERP Paradox

A measurement is made in the near field at some distance, D, from an antenna. It is determined that the measured power at that distance is P. By the inverse-square law (in the near field) it can be calculated that at distance D/2 the power would be 4P. Make D/2 a new point of measurement and repeat the experiment. You can see that for each D there exists a point D/2 that lies halfway to the antenna. At each successive selection of the measurement point the power increases by a factor of 4. As you approach the antenna itself the power will become infinite.

The paradox will be exposed in the discussions that follow so don't spend too much time struggling with it. If you would like a test to see whether or not someone is really "in the know" about all things electromagnetic, throw Bardwell's Paradox at them and see how they respond. The "answer" to Bardwell's Paradox (and Zeno's) is given in Appendix A and will be best appreciated after reading the remainder of the text.

How, then, shall we appreciate the nuances of electromagnetic field propagation and behavior? If we could see (with our eyes) what the RF field propagating from an 802.11 access point looked like then we would be better prepared to study it. Unfortunately, we can only use tools like spectrum analyzers and 802.11 protocol analyzers to gain indirect insight into the electromagnetic vibrations in the air around us.

Section 3: The Electromagnetic Spectrum

Your eyes are remarkable organs. They have a physiological, biological mechanism that is highly sensitive to electromagnetic radiation. Of course, the radiation that your eye detects is in the very high frequency range (short wavelength range) that we call visible light. If your eyes could actually see in the wavelengths below visible light you could look at the electromagnetic field in the air. It would have gradations of “brighter” and “dimmer” where the signal was stronger or weaker, and, in some strange sense, a different “color” for different frequencies. Perhaps 802.11 Channel 1 would be “red-like” and Channel 6 would be more “blue-like” and an area of no signal coverage would be “dimmer”.

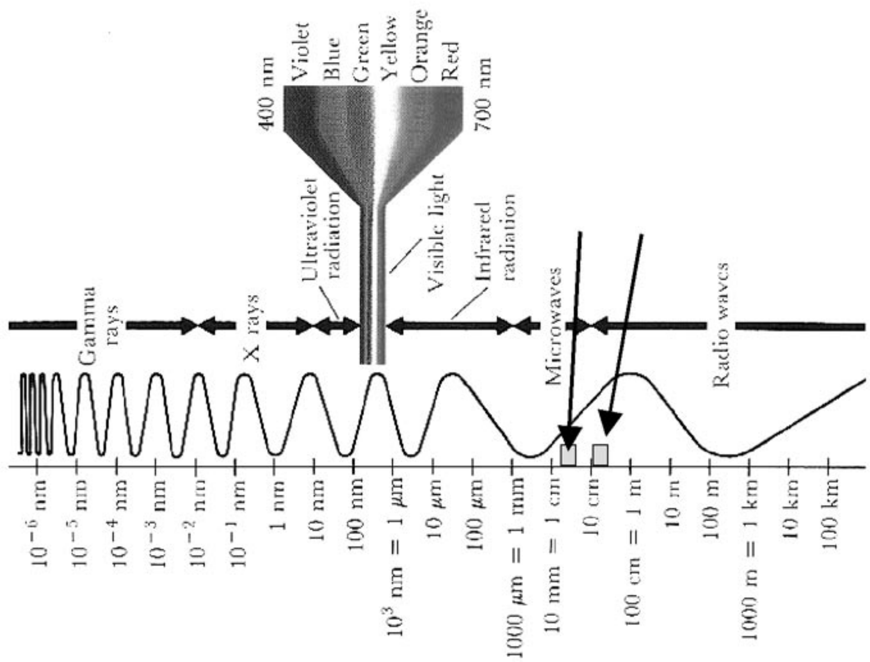


Figure 3.1 The Electromagnetic Spectrum

Unfortunately (or, perhaps, fortunately from the standpoint of human evolution) you can’t use your eyes to see the expansion, rotation, and various disruptions and attenuations in the entire electromagnetic spectrum that surrounds you. When you see a Ferris wheel at a carnival your brain interprets the electromagnetic radiation patterns that fall on the back of your eye as a large, circular sort-of thing with seats, going around and around. When you’re in a room with only red lights, your brain interprets the color green as if it were black because your friend’s green shirt doesn’t have any green light that can be reflected back to your eye. The bottom line is that when electromagnetic radiation in the visible spectrum is transmitted, reflected, absorbed, refracted, or diffracted, the resulting pattern can be observed by your eye and interpreted in some manner by your brain. You don’t have the same luxury with the 2.4 GHz or 5.8 GHz transmissions in an 802.11 wireless network. Consequently we’re going to have to use our imagination to try to picture what the radiation patterns might be in a particular room, from a particular client machine, or from an access point. Perhaps the signal is being reflected off the metal Venetian blinds or filing cabinets. Perhaps the large water fountain in the courtyard is absorbing much of the transmitted energy. Perhaps that big advertising billboard outside the window is acting as a point of signal diffraction and is causing lost connectivity across the street. How can you know when you can’t see the radiation?

The Shape of the Electromagnetic Field

Let's try to build a mental picture of what the radiation pattern might look like if we could see radiation in the 802.11 frequencies. To begin we will define an imaginary radiating object, a perfect radiator. From this perfect radiator the electromagnetic field will propagate outwards in all directions equally. This theoretical object is called an *isotropic radiator*. Seen in cross-section the isotropic radiator would appear as a perfect circle, with the radiating element being a point in the middle and the electromagnetic field expanding equally in all directions. In three-dimensions the electromagnetic field would look like a sphere.

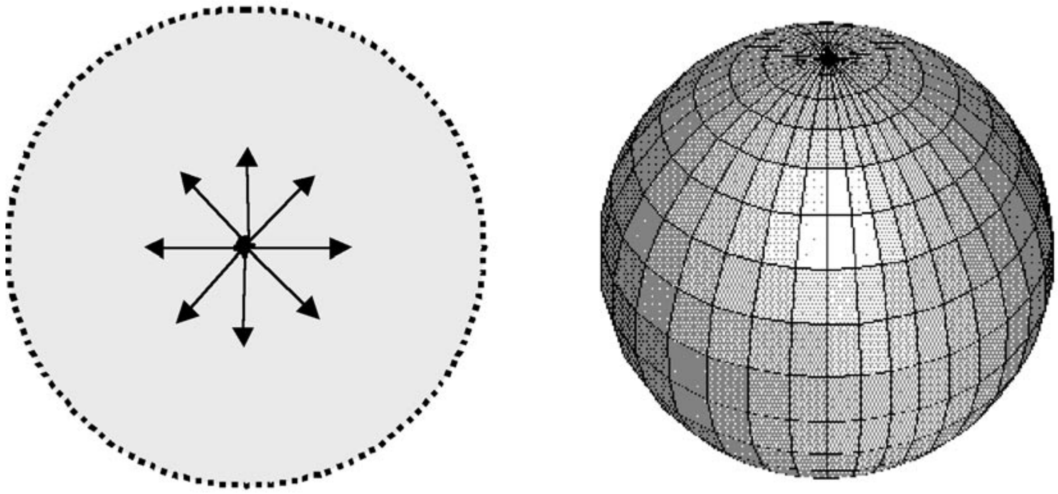


Figure 3.2 The Spherical Radiation Pattern of a Theoretical Isotropic Radiator

A real antenna cannot have an isotropic radiation pattern. As was discussed earlier, the electromagnetic field propagates in a direction that's perpendicular to the radiating wire so, for a vertical antenna, the field isn't going to propagate upwards and downwards in the same way it propagates to the sides. This picture would more resemble a doughnut placed down on the antenna, as shown below.

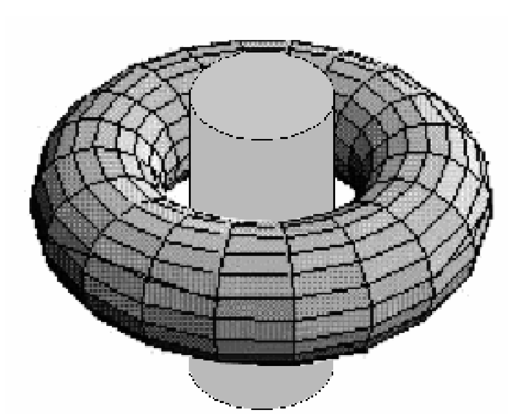


Figure 3.3 The Doughnut-Shape of the Electromagnetic Radiation Pattern

This picture describes the general shape of the field surrounding an antenna type known as a *dipole*. Simply put, a dipole antenna has two ends with differing electric polarities. The typical antennae used by an 802.11 access point are normally dipoles. The actual field isn't exactly as shown above, but this should give you a starting point for understanding how a simple dipole radiator will behave.

Realize that the reason there isn't any field directly above or below the radiating element is because the field expands outwards, at 900 from the element. If it weren't for the fact that the propagating signal is bouncing around inside a building (or, to a lesser degree, off the ground and other objects outdoors) you would find that if your 802.11-enabled notebook computer were directly above, or directly below a transmitting antenna there would be no signal. Someday, if you find yourself floating around in outer space, with no spaceship nearby (which would probably be a bad thing!) you'll be ready to troubleshoot the lack of connectivity between two 802.11 devices floating along with you. In free space, with no reflective objects to interfere with the theoretical propagation pattern of a dipole antenna, the absence of signal above and below the radiating element would be significant. Some consideration may be given to this characteristic of propagation in practical settings, but the chances of being noticeably impacted by the zone of reduced signal strength are insignificant.

Particles and Waves

Is the electromagnetic field that propagates from an access point to an 802.11 wireless client (or, light, for that matter) some kind of little particles shooting out, or is it some kind of wave-like stuff, like water waves spreading out after a stone is thrown into a calm pool? The answer is "both." This mystery has been challenging scientists for over one hundred years and elements of the mystery were part of the discovery of how electromagnetic fields are affected when they encounter an obstacle in their propagation path. Since there may be obstacles between an 802.11 transmitter and receiver the mystery becomes our own potential enigma.

Historically the argument about whether an electromagnetic field was made up of particles or waves went back and forth with a number of fascinating "in-betweens" over the years. The essential contradiction that puzzled people was the fact that Newton's laws of motion seemed to apply to light (as if light were made up of little particles) but light interference (like the splitting into color bands in a rainbow or glass prism) follows the rules of geometric optics (as if light were made up of waves of energy). James Maxwell developed equations that treated electromagnetic fields as if they were made up of wave-like disturbances of the "aether" which permeated all of space. It was Albert Einstein who showed that no medium ("Ether") was required for the transmission of electromagnetic waves. In the mid-1900's Richard Feynman built on the recognition that the electromagnetic field did, in fact, consist of energy quanta (tiny energy packets) and developed the discipline known as Quantum Electrodynamics (QED). Fortunately, as far as the real-world realm of 802.11 wireless networking is concerned, Maxwell's basic laws apply completely and there's no need to think about quantized decomposition or other exotic theories. Suffice it to say however, at the smallest distances (sub-atomic distances) Maxwell's laws don't exactly explain all of the particle-like behavior that can be observed by physicists. These two viewpoints were congealed in the strange world of Quantum Mechanics, which offers theories in an attempt to explain how particles (like photons and electrons) can act as if they were in two places at the same time, interfering with themselves and defying direct experimental evidence of their location and energy state. This mind-boggling characteristic of wave-particle duality is called *quantum superposition* and the underlying principle centers around the idea that a particle "thinks" about all possible paths that it could take going from Point A to Point B and a *probability amplitude* can be calculated indicating the chances of finding the particle at any particular point. Erwin Schrödinger conjectured on the improbability of these characteristics when he suggested his famous thought experiment with a cat in a box. The fate of the cat hinged on the quantum behavior of the experiment's imaginary apparatus. The conclusion was that the cat would have to be both dead and alive at the same time, until the box was opened. He did, however, develop an equation that predicted the probability amplitudes for determining a particle's location and his equation (the Schrödinger Wave Equation) is a fundamental part of engineering practice and study today.

To gain an idea of just how strange this idea really is, let's digress for a moment and consider a beam of light being aimed at an angle towards a mirror as shown below in Figure 3.4. You probably learned in high school physics that the "angle of incidence equals the angle of reflection," and, in terms of classical physics, this is true.

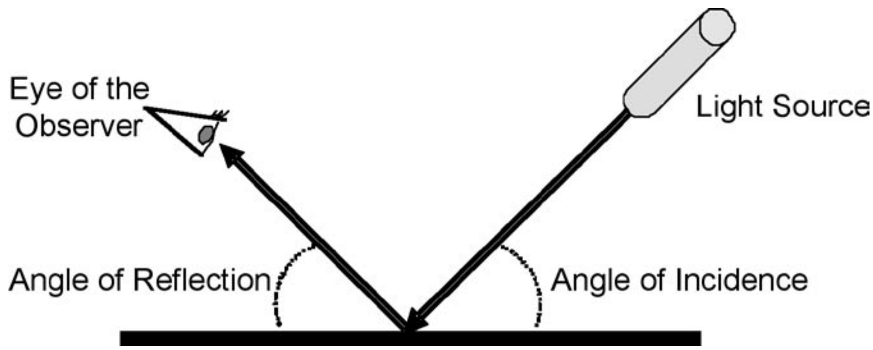


Figure 3.4 A Beam of Light Reflecting From the Surface of a Mirror

Augustin Fresnel (pronounced "Fra-nell") was a French scientist and mathematician who, in 1816, published research based on his belief that light acted like a wave (as opposed to the "corpuscular" or particle view that was predominant at the time). In fact, he was generally unaware that the particle view of light was held as being "correct" in his day and so he freely pursued what was, at the time, an unpopular avenue of research. His research focused on how a beam of light was diffracted as it passed through, or was reflected from, a very narrow slit, or series of closely spaced slits. He found that "... elementary waves arise at every point along the arc of the wave front passing the diffracter and mutually interfere." He went on to say "The problem was to determine the resultant vibration produced by all the wavelets reaching any point behind the diffracter. The mathematical difficulties were formidable, and a solution was to require many months of effort."

His work served as the foundation for creating magnifying lenses (Fresnel lenses) that took advantage of what would later be studied to include the possibility of quantum mechanical effects. If a surface is created that reflects or diffracts light at some points, and completely blocks it at other points, the probability amplitude for the light beam reflecting with an angle of reflection equal to the angle of incidence is reduced. The light can still go from the source to the eye of the observer, but it takes an alternate path (since the "best" paths are blocked by the physical construction of the reflecting or diffracting surface. It seems impossible that this should be the case, but it is not only mathematically demonstrable and known as a fact by physicists, but the little flat "book magnifiers" that you might find for sale in a book store are examples of the phenomenon. The next time you see one of those flat, plastic sheets that you lay on a page and the text is magnified, you're looking at a Fresnel lens in action. Figure 3.5 (below) shows how a beam of light will act when the "best" point for getting from the source to the eye of the observer is made non-reflective (or non-diffractive, in the case of a transparent lens).

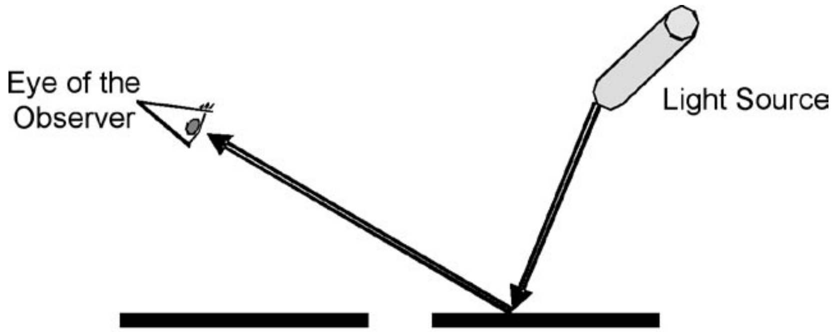


Figure 3.5 A Beam of Light Manifesting Fresnel Diffraction

As with many complicated examples that have been suggested, this is a dramatic oversimplification of the characteristics of Fresnel reflection or diffraction, but it is fundamentally accurate. There is no question that some readers will consider the diagram above and conclude “No way… that’s just plain wrong.” In fact, a study of the nature of electromagnetic waves, and the work of Fresnel, Huygens, and Poisson, and Laplace, will confirm that this is really representative of reality.

In the realm of 802.11 wireless networking the nature of electromagnetic waves makes it realistically impossible to perfectly predict how a particular implementation of access points, specialty antennae, and mobile clients will interact. For this reason, the wireless network designer should never assume that a particular design would work properly at the extremes of specified values. Consider Figure 3.6 below. If a vendor specifies that a highly directional antenna has a range of 5 miles, you shouldn’t quote the equipment for a 15-mile span assuming that you’ll only need 2 repeaters with 6 antennae without performing a site survey first to experimentally determine whether or not other influences reduce the range.

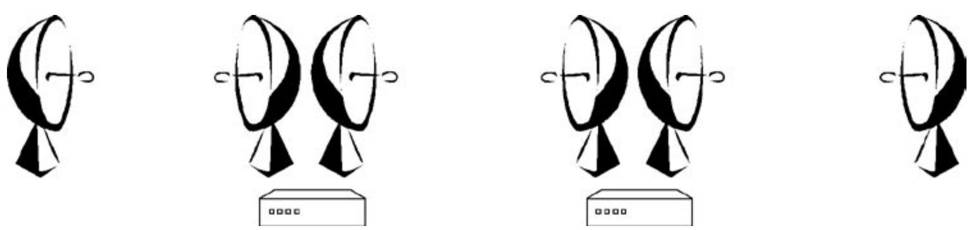


Figure 3.6 A 15-mile Span Using 6 Antennae and 2 Repeaters

By the same reasoning, just because an access point may provide coverage for 50 clients within a 200-foot radius in one building doesn’t mean it’s going to work that way in a building with different construction or different contents. In fact, it might not work the same way when the relative humidity goes up or when the galactic center is predominant in the night sky!

Sometimes even unexpected environmental influences impact communication in 802.11 networks. You may, or may not know that our sun goes through an 11-year cycle of sunspot activity. When the sunspot cycle is at its peak violent electrical storms on the surface of the sun bombard the Earth with massive amounts of radiation. These storms are known to interfere with many different RF communication systems including broadcast and 2-way radio, cellular phones, and our own 802.11 wireless networks. The behavior of the electromagnetic field in an 802.11 network, then, is acted on by many influences. Some of these are the result of the transmitted signal itself, and others result from environmental factors.

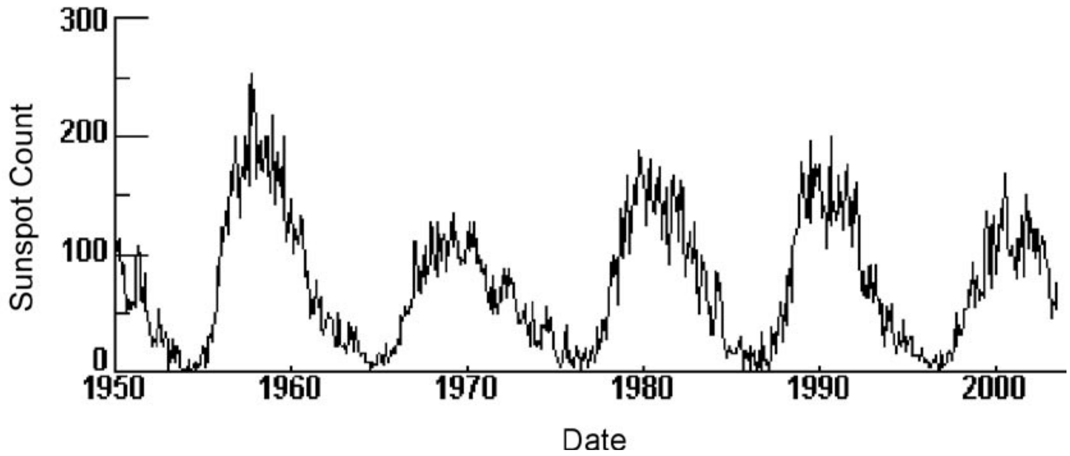


Figure 3.7 Monthly Sunspot Activity Since 1950

Since the earliest consideration was given to the nature of the electromagnetic field and how it behaves there's been discussion concerning whether the field is made up of some type of particles or whether it's actually some kind of wave. Some considerations regarding electromagnetic signal propagation are more consistent with the particle view of energy transmission, while others are more consistent with the wave view. To begin exploring these considerations, it's good to start with the most basic (and historically earlier) detailed explanations of the electromagnetic field presented by James Maxwell.

The Electromotive Force

When speaking about electromagnetism there are some fundamental quantities that are considered, calculated, estimated, and measured. When we talk about a transmitter sending some energy out into space we're ultimately talking about some electrons that were moving around in the radiating element and a magnetic and electrical field moving outward from the radiator. Electrons are in motion and the resulting electromagnetic field is propagating. The field is ultimately going to do the work of moving the electrons in the receiving antenna.

While it's electrons (particles with mass) that are moving in the electrically conducting element of an antenna it's an electric and magnetic field (forces without accompanying mass) that move in the field pattern surrounding the antenna. It was Albert Einstein who equated mass and energy. In fact, we can consider the implications of Einstein's work in the context of field propagation. Moving objects with mass have momentum. Isaac Newton pointed out that an object in motion would remain in motion unless acted on by some outside force. He described the quality of *momentum*. When an object's mass or its velocity is increased it becomes harder to change the velocity or direction of the resulting motion. It turns out those equations considering things without mass (like electromagnetic radiation) can involve something similar to momentum as well, even if the thing that's moving doesn't actually have mass. The characteristic that's similar to momentum is called momentum density and it manifests itself with a vector quantity called the vector potential of the field. Thank you Albert.

In the discussions of electrostatics and electromagnetism that follow, only the simplest features will be involved. Armed with some solid basic math and a hint of calculus, the reader should have few problems gleaning insight from the text.

The study of field propagation and electromagnetic theory hinges on three basic concepts. The simplest is that of the existence of an electrical charge with either a positive or negative value. Quantum physicists and theoreticians can explore even this seemingly simple concept throughout an entire career lifetime. We will not be exploring it at that depth! The second principle is called Coulomb's law and it stipulates that the electrostatic force between two charged particles is proportional to the product of their charges (q1 multiplied by q2 where q is the force of the charge) and is inversely proportional to the square of the distance between the two charges. The force acts along a line between the two charges. The third basic concept is that charges always influence each other. That is, if a force is exerted on a charge by a second charge (force F1) then it follows that a force is also exerted on the second charge by the first one (F2). This principle shows that the total resultant force between the two charges is the vector sum of F1 and F2.

An 802.11 transmitter radiates an electromagnetic field into the space around it. Aspects of the environment affect the field. There is an alternating electric current (positive and negative charge) being impressed on the metal radiating element in the transmitting antenna. The expanding electromagnetic field is an undulating energy field with both electric and magnetic characteristics. At the moment current is first applied to the antenna the electromagnetic field expands outwards at the speed of light. While the signal is being applied to the antenna the field energy ebbs and flows in a wavelike manner. These waves may be considered in their entirety by applying equations involving advanced calculus, specifically, the functions that allow an infinite set of values to be integrated over a surface area or in a 3-dimensional volume. The challenge is to gain a general knowledge of how all of these theoretical aspects of electromagnetism work and, in many cases, the only way to share that appreciation is through mathematical expression. Every attempt has been made to minimize any frightening aspects of mathematics or calculus that may creep into the discussions to follow.

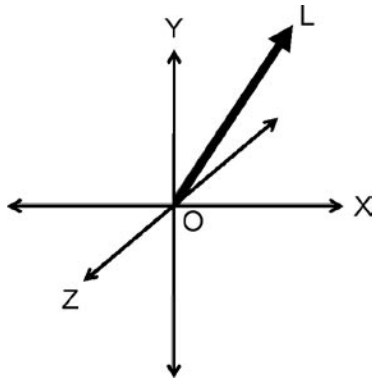
At points that are very close (less than 1-foot for 802.11 networks) to an antenna the electromotive force (the "pushing" force of the field) exerts itself to some degree horizontally (across the x-axis), to some degree vertically (up and down on the y-axis), and to some degree to the front and back (along the z-axis). Electromotive force (represented by the symbol "E") is, then, a vector quantity, having both magnitude and direction. The voltage in each direction is represented, by convention, using the symbols Ex, Ey, and Ez for the magnitude in each direction, x, y, and z. E, therefore, is the vector sum of Ex, Ey, and Ez. The variable "E" is only one of a number of metrics that are part of the study of electromagnetism. By gaining a general appreciation for a more complete set of fundamental metrics the character of the electromagnetic field can be more fully understood.

Scalar and Vector Measurement Metrics

When a child learns to read they preface their efforts by learning the alphabet. It's only when they comprehend the individual letters that they can recognize organized groups of letters as words, then group those into sentences, and ultimately appreciate great works of literature. The study of electromagnetic theory, too, begins by learning about a basic group of symbols and what each of these symbols means. The symbols are variables that make up mathematical equations which describe the behavior of electromagnetic fields. The mathematical languages that are used include algebra, trigonometry, and calculus. Calculus has sub-dialects that apply to different situations. PhD candidates in mathematics, physics, and engineering may be fluent in these mathematical languages. They, then, can read the exotic equations that describe electromagnetism and catch a glimmer of its mysterious character. We, however, will have to be content to struggle through only the elementary steps of mathematical representation. The goal will be to grasp the application of enough algebra, trigonometry, and calculus to get an appreciation for the complexities and beauty of electromagnetic field theory.

Nature manifests itself with a multitude of measurable characteristics. We can eat too many calories (a unit of heat) and then we gain weight (a unit of gravitational attraction). Our refrigerator keeps the milk at 40OF (units of temperature) through the conversion of watts of electricity (units of power) into torque (units of rotational force) and heat (units of molecular motion) in the compressor motor that is at the heart of the cooling system. The total number of measurement units in the whole of the sciences is sufficiently vast to exceed any person's capacity for remembering them all. Like the letters of a child's alphabet, it's the variables that make up equations that serve as the first step in comprehending electromagnetism.

Measurements like weight, speed, charge, and temperature are called *scalar* quantities. Scalar quantities simply represent the magnitude something (i.e. gravitational attraction, distance per unit time, number of electrons, or excitation of molecules). When a quantity includes both a magnitude and a direction of action, it's called a *vector* quantity.



On the left is a diagram showing the x-axis, y-axis, and z-axis of a 3-dimensional set of coordinate axes. The line from the origin (O) to a point located in space (L) may be considered a representation of a vector. The line has a length and it has a direction both of which can be determined from the location of point L relative to the x-, y-, and z-axis. This point would be specified as L(x,y,z) with the displacement on the x-, y-, and z-axis provided.

Adding two vectors is straightforward. Consider a hiker who started at their camp and walked 3 miles East. Then they walked 2 miles North. The two legs of the hike are vectors because they both have magnitude and direction. You can visualize how, on the map below, the two legs could be plotted and the hiker's position determined. This is the operation of vector addition on a flat plane (2 dimensions).

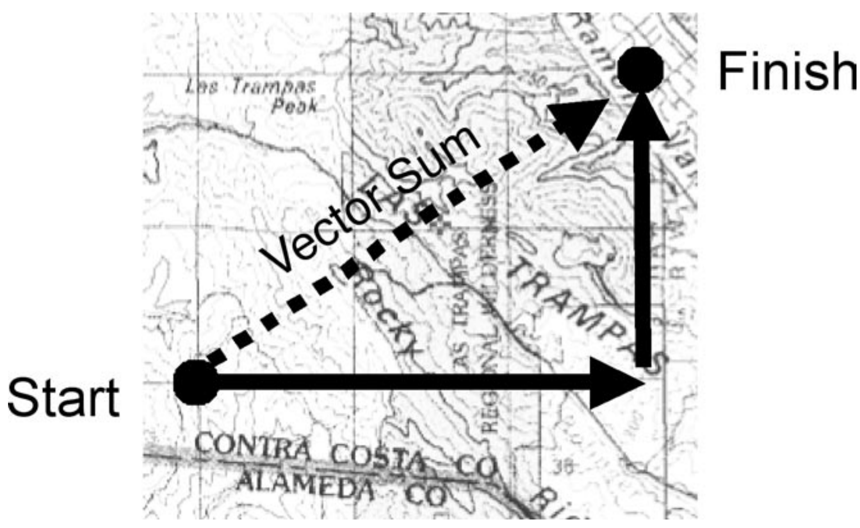


Figure 3.8 Hiking in the Las Trampas Wildlife Refuge

The vector sum for the two legs of the hiker's journey would be the dashed line in Figure 3.8 (above). The direction of the line could be calculated by the trigonometric Tangent function (tan) and the length of the line could then be determined using either the sine (SIN) or cosine (COS) function. These functions are explained in Appendix B.

Measuring the Characteristics of the Electromagnetic Field

An electromagnetic field manifests a number of different characteristics and interacts with the medium through which it propagates in a variety of ways. While it's beyond the scope of this paper to discuss the details and ramifications of each type of electromagnetic interaction, it is appropriate to explain the fundamental characteristics on the basis of their measurable metrics. In this way the complexity of the field itself can be better appreciated.

Many of the metrics used to describe an electromagnetic field are vector quantities. Although a single letter symbol is used for the metric ("E" for example) it must be recognized that there are three separate components (x, y, and z) that make up the quantity. Mathematicians have developed the convention of representing the three components of a vector as a single letter symbol and it's understood that when this symbol appears in an equation there are multiple parts to the calculation. You should note that the measurement of permittivity and permeability (discussed below) are not vector quantities but, rather, are simply constants of proportionality that represent physical characteristics of matter. Electric Field Strength (E) and Magnetic Field Strength (H) are the two fundamental field metrics.

Electric Field Strength Vector Symbol is: E

E is the measure of volts per meter. Since E in this case is a vector quantity it takes into consideration the fact that the force has components in the x, y, and z direction. This is different from the V (voltage, also commonly represented by E) in an electrician's use of Ohm's law (which doesn't involve vector quantities) since the electrician is concerned with current flow in a wire with, essentially, only one direction to consider. Computing V in a wire doesn't demand the use of the x or y directions, only the z direction (the length of the wire). Computing E in free-space is a 3-dimensional exercise.

Magnetic Field Strength Vector Symbol is: H

A magnetic force results from the movement of charged particles (in a dipole radiator antenna, for example). The intensity of this magnetic force is measured in units of "Newton seconds per Coulomb meter" or Amps per meter. Magnetic force, H, is thought of as the *magnetizing force*. It is this force that is amplified when, for example, coils of wire are wrapped around a central core with each individual wrap contributing its own field force to the overall magnetic force of the wrapped coil. The unit of measure for the Magnetic Force is the Ampere-turn. This magnetizing force is what affects substances placed near to the radiating element with the affected substance taking on magnetic properties as a result of the magnetic force. The amount of magnetism induced into a substance by the magnetizing force (H) is called flux density and is represented by "B" described below.

H is measured in Webers per square meter. Remember that a moving electrical current produces a magnetic field, and a moving magnetic field induces an electric current in a conductor. When the intensity of the magnetic flux crossing a conductor is sufficient to induce one volt in one second the flux density is 1 weber. As it turns out, 1 single weber is a very small unit of practical measurement so an alternative unit, called a Tesla, is used in actual calculations. The Tesla is defined as the density of a magnetic field such that a conductor carrying one ampere at right angles to the field has a force of one Newton per meter acting on it.

The electric and magnetic forces do not remain static in an electromagnetic field. They are oscillating, forming a propagating wavefront. The oscillations do not occur without some opposition, however. Nature, to a greater or lesser extent, doesn't want these forces to change with unrestricted ease and so they must "push" in order to oscillate. The degree to which an electric field finds it either easy or hard to change is called permittivity and, for the magnetic field, this characteristic is called permeability. Notice that these metrics are not vector quantities and when they operate on a vector the calculation involves applying the metric singly to each component of the vector.

Permittivity in a Medium Symbol: ε

Permittivity is the characteristic of matter that refers to the proportion between the electric displacement and the electric field intensity. The symbol ε (epsilon). In a vacuum ε is a constant equal to 8.85 X 10^{-12} Farad/meter (F/m) and represented by ε_0 (pronounced "epsilon naught"). The Farad is a unit of capacitance and is measured in units of seconds to the fourth power per ampere squared per kilogram per meter squared (s^4A^2/kg/m). What permeability (μ) is to the magnetic field, permittivity (ε) is to the electric field. It's a measure of how easily the electric field is amplified when acted on by some particular electric field strength. The permittivity of a medium is the ratio of a value called the *electric displacement* (D, explained below) over the electromotive force (ε = D/E).

Permeability in a Medium Symbol: μ

Permeability is the characteristic of matter that refers to the proportion between the magnetic flux density and the magnetic field strength in a given medium. The symbol μ_0 is used to represent the permeability in a vacuum and it's a constant equal to 1.257 X 10^{-6} Henry/meter (1.257 μH/m "microhenrys per meter"). One Henry is the equivalent of one kilogram meter per second squared per ampere squared. Considering the units ($Km/s^2/I^2$) it can be seen that one Henry of permeability equates to the movement of one kilogram through a distance of one meter in one second when acted on by one amp. Permeability, then, is a measure of how easily the magnetic flux density changes when acted on by some particular magnetic field strength. Permeability in general is symbolized using μ without the zero subscript. The general permeability of any particular environment is affected by the walls, furniture, metal filing cabinets, trees, or whatever else might be present.

The concept of flux density is worth considering in more detail. Although an electromagnetic field is massless it has, nonetheless, areas of greater and lesser magnetic flux density. It can be compared to a medium with mass, such as a spoonful of lard in a frying pan which, as it melts, has areas of greater and lesser density (less melted and more melted). As a consequence of flux density, the laws of classical physics relating to momentum can be applied to an electromagnetic field as a whole. The field may be thought of as if it had a center of gravity which conforms to the basic properties surrounding F=ma (force equals mass times acceleration).

When the electromotive force (E) is considered with regard to the permittivity of the medium (ε) the result is called the *electric displacement* (D, where D = εE). In a similar manner, when the magnetic force (H) is considered with regard to the permeability of the medium, the result is called the *magnetic flux density* (B, where B = μH).

Electric Displacement Vector Symbol: D

D is the density of electric charge per unit area that would be displaced across a conductor placed in an electric field. The displacement (change in the electric field) is in the direction of the E vector quantity, proportional to the permittivity of the medium such that $D = \varepsilon E$. D is measured in Coulombs/m2. In many respects the electric displacement (D) relates to the electromotive force (E) in a manner that is similar to how the magnetic flux density (B) relates to the magnetic force (H).

Magnetic Flux Density Vector Symbol is: B

A magnetic field is characterized by its ability to induce magnetism into a substance or object over which the field extends. The amount of magnetism (magnetic induction) that is induced into a substance as a result of the magnetizing force (H) is the Magnetic Flux Density and is represented by the vector symbol "B". The intensity of B is affected by the intensity of H and a characteristic of the substance being magnetized called permeability. The value of B is equal to the permeability multiplied by the value of H ($B=\mu H$, where μ is the permeability).

Current Density Vector Symbol: J

J is measured in units of Amps/meter^2 and may be thought of as a position dependent density of electric current. Since moving current is a source of magnetic fields, B is related to J in proportion to the permeability of the space in which J exists (μJ).

Vector Potential for the Field Vector Symbol: A

The vector potential for a field is the simplest of concepts, and yet one of the most challenging to discuss. To begin understanding the vector potential, it's necessary to understand *Gauss's theorem*, also called the *divergence theorem*. The divergence theorem expresses that unless matter could be created or destroyed, the density within a region of space can change only if something flows either into or out through the region's boundary. This statement is sometimes simply called the *principle of continuity*. The divergence of a vector field is the rate at which flux density exits from the region. A magnetic field (B) has a potential (in the x-, y-, and z- directions) to diverge. The potential to diverge is called the *vector potential* for a field.

Electromagnetic Momentum Vector Symbol: P

Momentum, in classical Newtonian physics, is defined as "mass times velocity." Force and energy are interrelated to momentum inasmuch as they are interrelated to velocity and mass. Electromagnetic momentum is, in the largest sense, a very complex subject since the photons that carry the energy of the electromagnetic field have no mass. There is, however, a characteristic of an electromagnetic field called *momentum density* that is the result of movement of the field in a manner similar to how momentum is the result of the movement of a mass. If an electromagnetic field is being propagated from a radiating antenna and suddenly the source current to the antenna is shut off (or changes polarity) there is a brief continuation of the outward radiation due to the electromagnetic momentum of the field itself. "P" is the momentum of the field and is measured in Amp meters squared (Am^2).

In the 1870's Sir William Crookes invented a device consisting of a sealed glass tube containing electrodes in a vacuum. He was able to demonstrate that the electrons flowing between the electrodes exerted a pressure on anyobject placed in their path hence showing that these particles had momentum. This same principle is extended to photons of light in the radiometer, an evacuated glass bulb in which black and white vanes are suspended on a pivot. When light strikes the vanes the photons yield their momentum and the vanes spin. Even in the absence of mass (as with photons) the characteristics of classical momentum are observed.

Radiometer

Differentiation of Functions with One Independent Variable

All of the fundamental characteristics described previously are interrelated. James Maxwell set down these relationships in the form of a set of equations known simply as the "Maxwell Wave Equations". We'll consider the wave equation that relates A to B to convey the types of interrelationships that exist between field characteristics and to continue developing a mental model of how electromagnetic fields behave.

The power force delivered by a moving electron is, in classic electrodynamics, considered to be a function of the acceleration of the particle. Acceleration of charges results in the creation of the electromagnetic field which then radiates away from an antenna. The field, then, carries energy away from the antenna and the rate of change of the field directly relates to the density of the magnetic flux. There is a specific relationship between the vector potential (A) and the magnetic flux density (B).

Compare an accelerating electric charge to a race car at a drag race. The car is going to go from zero to 300 miles per hour in just under 5 seconds. Assume we could measure the distance that the car had traveled (position) at each moment of the race (time). A graph of the distance versus the time during the time that the car was accelerating might look something like Figure 3.9 below.

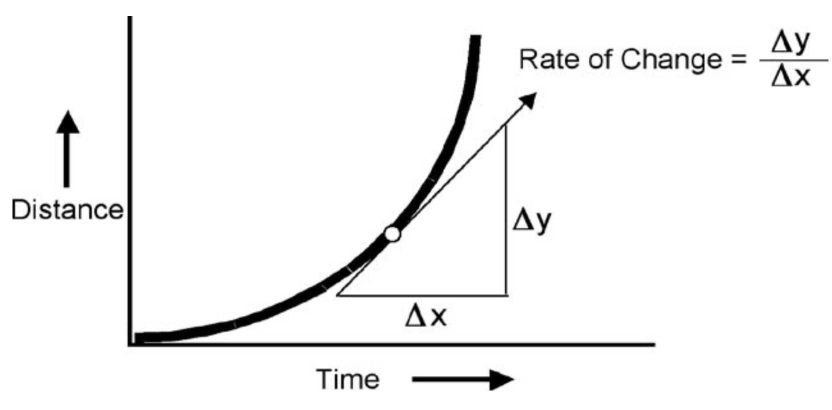


Figure 3.9 Position Versus Time and the Rate of Change

Consider the state of the car's motion at the point shown on the curve. If a line is drawn through the point tangent to the curve, it represents the speed of the car at that point. The arrow on the line that's tangent to the curve represents the velocity (miles per hour) at the point shown. The slope of the line ($\Delta y/\Delta x$) is the rate in units of distance over time. A mathematician says "Velocity is the first derivative of position with respect to time" and the equation would be written as follows.

$$\frac{dk}{dt} f(k) = \frac{\Delta y}{\Delta x}$$

$$\frac{dk}{dt}$$

This notation means, "differentiate k with respect to t"

Figure 3.10 The Notation for Differentiation

The equation in Figure 3.10 says that there is a function that operates on a variable "k" and that function is going to be differentiated so that for any value of t the rate at which the value of k is changing in the function can be determined. The "dk/dt" notation is not a fraction or ratio: it's a symbol that means "differentiate k with respect to t". Differentiation means, simply, find the slope of the line tangent to the curve of the function for the variable specified as it's varying relative to the "respect to" variable.

Given an equation for time and position the first derivative yields velocity (with the magnitude of velocity being speed). Now, speed itself may be changing and that, of course, is called acceleration. If an equation was presented that gave the function relating time and speed then the first derivative of that equation would be acceleration. This would then be the *second derivative* of the function relating time and position. The mathematician says "Acceleration is the second derivative of position with respect to time." The second derivative of a function may be indicated by putting a small number 2 in a superscript position above the "d" in the top part of the symbol, and above the variable in the bottom, as shown to the right. This does <u>not</u> mean that something is raised to the second power. The superscript 2 is simply a way to indicate that the second derivative of the function is being considered.

$$\frac{d^2k}{dt^2} f(k)$$

This mathematical abstraction continues. The change in the rate of acceleration may also be significant in some circumstances. For example, your coffee will spill out of the cup when you're driving if the rate of change of acceleration varies too dramatically and you can't tilt the cup accurately enough to compensate for the change in velocity. The ISO (International Standards Organization) Vibration and Shock vocabulary (ISO 2041) refers to this metric as "jerk" and it's the third derivative of position with respect to time. Trains are expected, for example, to keep jerk to less than 2 meters per second cubed for passenger comfort. The aerospace industry even has a jerkmeter that measures jerk. A superscript 3 with the "d"s in the derivative symbol would indicate that the third derivative is being considered.

And, not stopping with three, there's even a use for the fourth derivative of position with respect to time. This derivative doesn't have an official name, but it's been called jounce and snap. Jounce, the rate of change of jerk, is taken into consideration in the design of sophisticated devices like the Hubble space telescope. The fifth and six derivatives have been called crackle and pop respectively although these names are not often taken seriously.

Differentiation of Functions With More Than One Independent Variable

Consider the characteristics of the electromagnetic field as it's being energized by the moving charge in the antenna. The field is propagating outwards from the longitudinal axis of the antenna and it's rotating (as per the left-hand rule) around the antenna. The vector potential has x, y, and z components that are all changing. It's not, therefore, possible to simply differentiate a function containing Ax with respect to y (d/dy) without taking the fact that Az is changing and affecting the situation too. Here, then, are situations where differentiation must be performed on three variables (the three components of A, for example) all at the same time.

When multiple variables are present (as is the case with the x-, y-, and z-axis values of a vector quantity) the equations must be differentiated on the basis of each variable separately while the other variables are temporarily held constant. This operation is called *partial differentiation* and it's represented by the symbol ∂, which is a stylized, curved letter "d" (and not a strange letter of the Greek alphabet). A partial differential equation is an equation for an unknown function of more than one independent variable expressed in terms of a relation between the partial derivatives of the function.

One of the most basic of Maxwell's wave equations sets forth the relationship between vector potential (A) and the magnetic flux density (B) and serves as an example of how calculus, the language of electromagnetic field theory, represents the physical characteristics of the field.

Magnetic Flux Density (B) and the Vector Potential (A)

Maxwell described how each of the three vector components of B could be determined based on the components of A. He set forth the following description of the Magnetic Flux Density (B):

- Bx can be determined from AZ and AY
- By can be determined from AX and AZ
- Bz can be determined from AY and AX

We are now ready to express the most basic of the relationships put forth in the Maxwell Wave Equations. The three equations below show how the various partial derivatives of the components of A interrelate to produce the components of B. The cyclic operation of computing one, then another, and finally the other component of a vector in this manner is called the *cross product*. In this case we're calculating the cross product of the partial derivative of each of the vector components.

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \qquad B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \qquad B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}$$

Figure 3.11 Partial Differentiation to Compute the Components of B

The mathematical relationship between the three vector components in Figure 3.11 is called the "curl of A". Computing the curl of a vector means processing each of the components of each of the vectors in the manner shown. The group of mathematical operations expressed by the three equations shown above would be written using either of the two following notations:

$$B = \nabla \times A \qquad\qquad B = \text{curl } A$$

The upside-down triangle (∇) is called the "del" operator and refers to the partial differentiation that takes place between the components of the vectors. The notation on the left is read "B equals the derivative cross product of A" and on the right, simply "B equals curl A". Both notations mean the same thing.

This equation (using the curl operator) means that the magnetic flux density (B) has a rotational relationship to the vector potential (A). This is the basis for the famous Maxwell Equations in Electrodynamics. Maxwell's equations express the relationships between the fundamental quantities related to electrical charge and magnetic fields. The full set of equations includes two additional special vector operators *grad* (gradient) and *div* (divergence). As with curl, these two operators take a series of vectors, component-by-component, and perform operations on them. The div operator takes a vector field and represents how much the field is expanding or being created at each point. Divergence allows calculation of how much a field is moving in or out and curl calculates the rotational characteristics of the field. The grad operator takes individual scalar components and equates them to a vector field. The gradient of change over the field shows where the field is changing most rapidly and would give the direction of field movement.

There are various forms in which Maxwell's equations are presented, some using curl, div, and grad and others expanding the equations so that they can either be solved or so that other characteristics can be emphasized. The basic equations, in vector form, are shown below.

$$\nabla \cdot \mathbf{B} = 0 \qquad \nabla \mathbf{x} \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho \qquad \nabla \mathbf{x} \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{J}$$

Figure 3.12 Basic Maxwell Wave Equations in Vector Form

In the wave equations (Figure 3.12) you see the del operator (∇) and, on the right, the cross product symbol (X). On the left, the dot between the del and B or E does not mean multiply. It refers to an operation called the *dot product*. The dot product is an operation whereby a product is calculated between the length of two vectors and the cosine of the angle between them. Therefore, given two vectors X and Y: $X \cdot Y = xy \cos \theta$.

The electromagnetic field not only obeys the mathematical laws expressed in the field equations, it also manifests the strange mathematical consequences of the equations. One of these consequences relates to the variable t (time) in Maxwell's equations. By the laws of simple arithmetic it can be deduced that t could be either positive or negative and the equations would not break down. There are no equations that yield infinite or undefined results (like y=1/x when x=0) and there are no equations that are forced into the exclusive realm of imaginary numbers (the square root of negative one). Nothing strange happens arithmetically when t is a negative number. This aspect of the field equations gives rise to a time symmetric characteristic of electromagnetic fields called the *Reciprocity Theorem*, discussed next.

Section 4: Electromagnetic Field Propagation

There is a basic principle of antennae that is so unexpected (to the uninitiated student) that some people refuse to believe it's true the first time they hear it. The principle is called the Reciprocity Theorem and the consequences of this theorem are that (for the same input power level) if you can hear my transmission then I can hear yours. Where this becomes astounding is when I have a very elaborate, high gain, highly directive antenna that's capable of beaming my 100 mW 802.11 signal across 20 miles. My antenna might be illegal, but if I'm trying to hack into your network, antenna legality is probably low on my list of concerns. You, on the other hand, have only the little nub antenna sticking out of the side of your notebook computer on your PCMCIA NIC. You have no special antenna, but you do have the same 100 mW power input that I do. The time-symmetric nature of Maxwell's wave equations form the basis for the fact that the same qualities of my antenna that allow it to beam my 100 mW signal over to you also allow it to work in reverse, picking up your signals from the air. Maxwell's equations are *symmetric with respect to time*.

Time Symmetry and the Reciprocity Theorem

To understand the significance of time symmetric physical properties, consider a pool table with a white ball near one end and the black ball in the center. The white pool ball is accelerated by the force of impact of the cue stick and travels towards the center of the pool table. In the center, the white ball strikes the black 8-ball in a straight, center-to-center impact. The inertia of the white ball is transferred to the black ball and it is now accelerated away from the white ball, in a straight line, leaving the white ball stationary at the point of impact. If you were to make a movie of the two balls striking and then played the movie backwards, it would show exactly the same thing except now it would be the black ball that starred in the opening scene of the movie. If the mass, velocity, and other characteristics of the Amazing Pool Ball Adventure movie were represented through mathematical equations the equations, would not be time dependent. Time could run forward or backward (negative time) and the results would be identical.

Consider the spatial volume of an expanding electromagnetic field. All of the fundamental metrics are measurable in the field. They are interacting in accordance with Maxwell's equations. The field propagates away from the antenna and the volume and its surface area get larger. As an electromagnetic field propagates it ultimately expands outward into the universe where, perhaps thousands of years from now, some being in a distant galaxy will be reading your 802.11-transmitted email. The concept is, however, one of *transmission volume*. Now, from what volume of space can an antenna receive signals? It turns out that the shape (in 3-dimensional space) of the transmitted electromagnetic field also defines the shape of the volume from which the same antenna can receive signals in accordance with the rule called the *Rayleigh-Helmholtz reciprocity theorem*.

Baron Rayleigh of Essex, England, published a theory of light scattering in 1871 that was the first correct explanation for why the sky is blue. He earned the Nobel Prize in 1904 for the discovery of Argon gas. He donated the proceeds of his prize to the University of Cambridge where he became Chancellor in 1908. Rayleigh based some of his ideas on the work of Ferdinand Helmholtz who, after fighting in the Prussian army against Napoleon went on to invent the ophthalmoscope in 1851. Both men studied electromagnetic phenomena and both would develop theories that would become cornerstones of present day physics.

The Rayleigh-Helmholtz reciprocity theorem states the following:

> *If an electromagnetic force of some particular magnitude is applied to the terminals of antenna "A" and the received current is measured at some other antenna "B" then an equal current (in both amplitude and phase) will be obtained at the terminals of antenna "A" if the same electromagnetic force is applied to the terminals of antenna "B".*

Perhaps the most unexpected implication of the reciprocity theorem is that a hacker who is trying to penetrate your wireless network doesn't have to be in the parking lot. They can be many miles away using a directional antenna. If they can transmit to you then they can hear your transmissions as well. Remember, of course, that the theorem demands that both stations have the same antenna input power.

To some readers the reciprocity theorem may be new. The implications of antenna reciprocity are far reaching and, if this is the first time you've encountered the concept, the implications may be too hard to accept without proof. In fact, not only is reciprocity demonstrable in the lab and in real-world installations, but the physicists of the world can provide mathematical proof that the theorem holds true. If the antennae and the space between them are replaced with a network of linear, passive, bilateral impedances, then the current through the network can be calculated in accordance with standard practices in electronics theory. Whether on paper, or in practice, given the same input power on both ends "If you can hear me, then I can hear you!"

Practical Considerations Related to Antenna Reciprocity

One practical application of the reciprocity theorem is in the determination of an antenna's reception pattern. Assume that you're working with 802.11 client machines having a 100 mW output power and using simple omnidirectional antennae. You're considering implementing some type of directional or other high-gain antenna. Perhaps you're thinking of using a panel antenna at the end of a long hallway, thinking that it could be used to communicate to offices on either side of the hallway. Will the panel antenna be able to receive signals from the client machines in the offices? You can answer this question by calculating (or considering) the radiation pattern that would result if you transmitted a 100 mW signal from the panel antenna. Since the power on antenna "A" (the panel) is the same as that which will be coming from antenna "B" (the client machine's NIC), the reciprocity theorem is applicable and, if the client would be able to hear the panel then the panel will be able to hear the client. The power flow is the same in either direction.

When you consider the reciprocity theorem in the context of an 802.11 access point some interesting observations can be made. The two antennae on an 802.11 access point implement antenna diversity. Only one of them is used to transmit at a time and the selection of which one should be used is based on which one previously received the stronger signal. Antenna diversity provides a way to partially compensate for the location of various transmitting sources and the possible types of signal effects that might be introduced by the environment (multi-path reflections, refraction, absorption, dispersion, signal interference, all of which will be discussed). The access point transmits back with the antenna that received the stronger signal.

Thinking about the doughnut-shaped transmission/reception volume that surrounds a dipole radiator, you'll see that the orientation (up, down, left, part-of-the-way left, etc). makes a tremendous difference for the orientation of the radiation field. With an antenna sticking straight up the doughnut lies parallel to the floor. With the antenna pointing straight out towards the back of the access point (antenna parallel to the floor) the doughnut is standing up like a giant automobile tire… great for the people on the floor above you and below you; great for the folks to the left and right of the access point; not so great for the folks on the same floor but in front of, or behind the access point. Have you ever seen someone configure his or her access point antennae to look like a 1960's television "rabbit ears" antenna? One antenna is angled to the left, and the other is angled to the right. Hopefully this isn't you. If it is, fix it before anyone finds out! Remembering how antenna diversity works, you can see that if the two antennae are at odd angles then the doughnut patterns for transmission and reception from the two antennae are tilted at opposing angles. This effectively defeats antenna diversity for most of the reception area. It's best to orient the two antennae straight up. The fact that access points have flexible antenna couplings is to allow the access point to be placed horizontally, vertically, or even upside-down on the ceiling while retaining the vertical orientation of the antennae. Figure 4.1 (below) shows a good and a bad orientation for the access point antennae based on an understanding of general dipole radiation patterns, antenna reciprocity, and 802.11 antenna diversity.

Figure 4.1 Correct and Incorrect 802.11 Access Point Antenna Orientation

Transmitters and Receivers with Different Power Levels

Another application of the Reciprocity Theorem is in the design of an 802.11 network in which the clients and the access points use different transmit power levels. Remember that antenna reciprocity only exists if both sides of a communication link use the same input power to their antennae. We've discussed the situation in which an access point with a highly directional antenna can push its signal out many miles to a client with a standard dipole antenna. Reciprocity tells us that the same characteristics that allow the directional antenna to get its signal to the client allow it to receive the client's omnidirectional transmission. This reciprocity only works, however, if both client and access point have the same (typically 100 mW) input power.

Consider what happens if the access point has a 200 mW input power to a standard omnidirectional dipole antenna and the client has a 100 mW input to an identical omnidirectional antenna. The client could find itself in a location where it could receive the signal from the access point but could not, with its 100 mW transmit power, send a response back to the access point. In effect, using an access point that is more powerful than a client "breaks" reciprocity and causes chaos in the network. Clients attempt to associate with access points that they will never reach.

Consider what happens when the client is more powerful than the access point. Let's assume that we've configured our access point to use only 60 mW of output power. This may be in an effort to minimize bleed-over of the signal from the access point into an adjacent space (like a parking lot). The 100 mW client is now going to communicate to the 60 mW access point. This situation could actually be beneficial. True, reciprocity is broken, but this time in a way that augments, rather than detracts, from overall network performance.

For the client to receive the signal from the lower-powered access point, the client is going to have to be closer than would have been the case with an equal-powered access point. The client's signal is going to cover more area (have a larger radius) than the access point's but, this doesn't confuse the access point or other clients. Moreover, since in a network using non-overlapping channels, the adjacent access points will be operating on non-interfering channels, so the expanded transmission area of the client (relative to the access points) isn't going to cause a problem.

Propagation of Electromagnetic Waves in Space

The simplistic doughnut picture of the electromagnetic field surrounding an antenna isn't really a true representation of what's happening in the air. To get a more realistic picture of what you would see if your eyes could see in the radio and microwave frequencies, we need to look more closely at how the electromagnetic waves are created and then propagated through space.

Figure 4.2 (below) shows that the physical form factor that is the "antenna" sticking up from, say, an 802.11 access point, is not the antenna element itself. The actual radiating element is inside the form factor and may consist of wound wires, etched areas on a circuit board, or it may use the form factor itself as a radiator. There are a wide variety of antennae, some of which will be discussed later in this paper, however a good place to start is with the simple, omnidirectional (all horizontal directions equally) dipole antenna. All single element dipole antennae are considered omnidirectional, however all omnidirectional antennae are not necessarily dipoles. One characteristic of a dipole radiator is that the length of the radiating wire is very short compared with the distance light travels in one oscillation period.

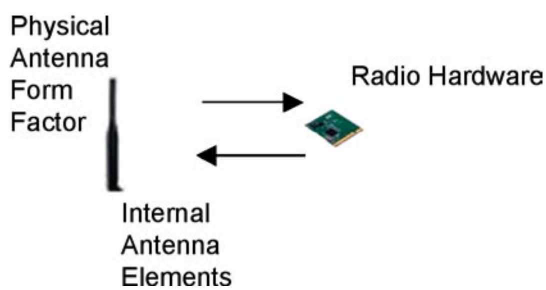


Figure 4.2 The Radiating Elements of a Dipole Antenna

Figure 4.3 (below) shows that a dipole antenna (as would be found on a typical 802.11 access point) consists of two separate radiating components with a common axis of orientation. As the radio hardware generates an oscillating signal (the 2.4 GHz carrier in 802.11b or 802.11g) the current is alternately pushed and pulled, in and out of each radiating component. This alternating electric current creates a magnetic field around the antenna.

The magnetic field grows in intensity as the magnetic forces move with the changing electric current and, as the signal travels through the dipole elements, a field emanates from the antenna. As the polarity of the driving signal changes the outward moving electromagnetic field reacts. Close to the antenna some of the transmitted energy returns to the driving elements of the antenna. Further away, the electromagnetic wavefront simply continue to propagate outwards. Figure 4.3 (below) shows the creation and movement of the wavefront resulting from the oscillations in the dipole elements.

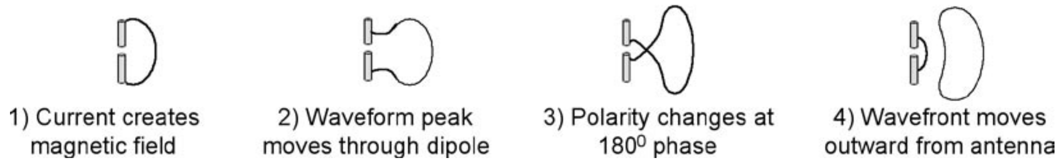


Figure 4.3 Wavefront Formation with a Dipole Radiator

Bear in mind that this diagram is only showing a cross section of the right-hand side of a three-dimensional "flattened doughnut" (torus) shaped field. It's the oscillating current in the radiating elements of the antenna that work together to create a magnetic field which, as the waveform rises to its maximum value, then falls back through the zero-point to its minimum value, causes the wave-in-space to get "squeezed off" the ends of the dipole. Figure 4.4 (below) is a reasonably accurate representation of what the electromagnetic field would look like surrounding a dipole radiator (specifically, a dipole radiator with an antenna length of $\lambda/2$, like an 802.11 "shortie" antenna roughly 6 or 7 cm long). Remember that this picture shows a vertical cross section of the "flattened doughnut" that surrounds the antenna. The horizontal double-headed arrow represents the perpendicular direction from the center of the antenna. If this were an antenna on an 802.11 access point it would be sticking straight up and the plane of maximum signal would be propagating outwards from it, parallel to the floor of the room.

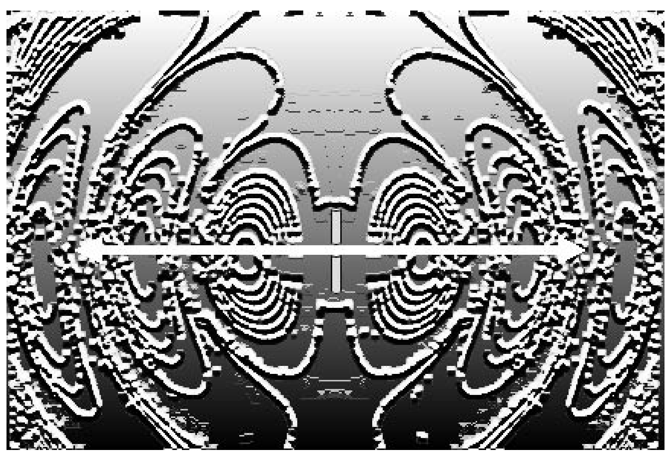


Figure 4.4 The Electromagnetic Field Surrounding a Dipole Antenna

It's important to understand that the field lines of force are produced as a result of the charge moving first up, then down, in the radiating element. The field moves outward with its strongest area being near to the plane perpendicular to the transmitting antenna (for a dipole). The lines of force are exerting their influence with the same oscillations as the original radiator. This implies that the physical orientation of a receiving antenna in the field impacts the degree of influence that the field has on it. The receiving antenna should be oriented in the same way as the transmitting antenna to make the field impart its maximum energy. Now, when you examine the figure above (3.4) you realize that there are places in space that lie above and below the dipole radiator in which the field lines are not even closely parallel to the radiating elements. This gives rise to the concept of antenna directionality. You can see that the closer you are to the horizontal plane that cuts the dipole into top and bottom vertical halves, the more directly the field will "wiggle" straight up and down.

A theoretical perfect radiator (one whose field radiates out in a completely spherical pattern) is called an *isotropic radiator* and it doesn't exist in the physical world. You can't build one. No matter what you may build there's going to be a point where one wire attaches, and a point where the other wire attaches. There's got to be some field orientation in any antenna that's constructed. A dipole radiator is one of the most reasonable approximations to the perfect isotropic radiator; the signal basically radiates out in all directions with, of course, a weaker area directly above and below.

All antennae have some particular field radiation pattern that is not spherical in the way an ideal isotropic antenna would be spherical. This deviation makes the field stronger in some directions, and weaker in others. This is the characteristic of *antenna directivity*. Physical characteristics of the radiating elements that make up an antenna unit cause the resulting electromagnetic wave to point more in some directions than in others.

Coupling and Re-radiation

By the Reciprocity Theorem we understand that the preceding description of dipole radiation can be thought of in reverse for dipole antenna reception. The electromagnetic field, varying in amplitude in the spatial volume surrounding the dipole receiver induces an electric current into the metal of the antenna. The antenna introduces electrical impedance and a voltage differential is created across the antenna terminals. The energy induced into the antenna can go somewhere. It travels through the wires from the antenna to the receiver circuitry where the signal is demodulated, amplified, and otherwise processed.

Consider now a situation in which the wires connecting the antenna to the receiver circuitry are cut. An accelerating charge in the magnetic field induces a current into the antenna but now there's no place for it to go. As the inducing field decelerates the antenna now radiates the energy back out into space, acting as a transmitter. This effect is called *re-radiation* and it occurs for any metal object into which a current is induced with no place to "drain off" to earth ground. While the effects of re-radiation might be detrimental if they unexpectedly occur because of a metal filing cabinet, they are used productively in the creation of directional antennae.

When two radiators are placed near to each other (within roughly 10-inches for 802.11 communications) they induce current into each other. In this situation the two radiators are said to be *coupled*. Coupling can be between two active radiators, each having a source of input power, or between an active radiator and a passive radiator (a metal rod acting only as a reradiator). In either case the electromagnetic field produced is the result of two (or more) separate radiating elements. This is the basis for directional antennae like the Yagi in which a single radiator induces current into reradiating elements that are designed to act as either reflecting components or as directing components.

Representing the Direction of Field Propagation

In 1884 a former student of James Maxwell named John Henry Poynting presented a paper entitled "Transfer of Energy in the Electromagnetic Field" at Mason Science College in Birmingham, England. He showed that the direction in which overall energy is flowing in a field could be expressed in a simple formula using the vector product of the electric and magnetic fields. It's indeed amusing that the equations to calculate the direction in which a propagating electromagnetic wave points were produced by a person named "Poynting." The Poynting vector is measured with dimensions of power per area (Watts/Meter^2). The equations for field intensity can be instructive with regard to building a mental picture of what's happening in that invisible (to your eyes) space between the transmitting and receiving antenna.

The Transverse Wavefront

When an antenna radiates an electromagnetic wave there is a spherical expansion (in general) to the expanding wavefront. Close to the transmitting antenna a receiving antenna would encounter a curved wavefront, since the sphere of expansion would grow outward to engulf the receiver. Farther away, however, the surface of the expanding wavefront would appear essentially flat to the short relative length of the receiving antenna as shown below in Figure 4.5. If r (the radius of the propagation sphere) is sufficiently large, the small section of the spherical wave that actually impacts the antenna may be considered as a plane. This small section of the surface of the expanding spherical wavefront is represented as a starting point (r, θ,ϕ) and then as the change in θ (dθ "delta theta" sometimes shown in equations as Δθ) and the change in ϕ (dϕ "delta rho"). The "d" (or Δ) means "change in". Figure 4.5 (below) shows an area on the surface of a spherical wavefront given as (r,θ,ϕ) and the two delta's that make up the vertical and horizontal extent of the area. Refer to Appendix C if you are unfamiliar with rectangular, cylindrical, and spherical coordinate system representation of vectors.

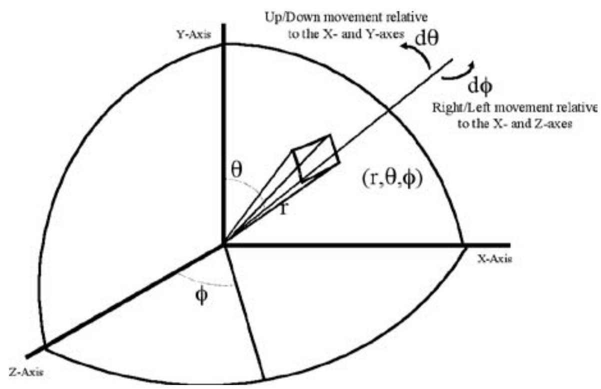


Figure 4.5 Surface Area Defined On the Spherical Wavefront

An 802.11 wireless NIC sees the expanding spherical wavefront as if it were a flat, planar wavefront because the length of the 802.11 antenna is so short compared to the curvature of the actual spherical expansion. This view is represented below in Figure 4.6.

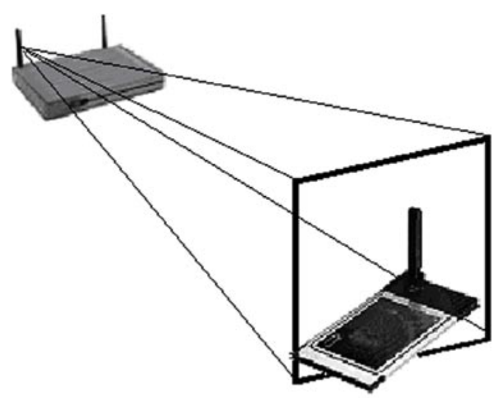


Figure 4.6 An 802.11 NIC Encounters a Flat, Planar Wavefront

The relationship between the perceived propagating planar wavefront and the interposed receiving antenna means that the electrical component of the wave and the magnetic component of the wave have a perpendicular relationship to each other. That is, we could think of the two aspects of wave propagation as being separate with the electrical field E vibrating up-and-down while the magnetic field (H) vibrating left-and-right. These two aspects of the wave are 180^O out of phase and their relative magnitude is dependent on the medium through which they propagate. This relationship is given as E/H = Z (where "Z" is the impedance of the medium which, in free space, is 377 Ohms).

The Electromagnetic Field Pattern

The varying electric field intensity as a function of radius and angle (θ,ϕ) is called a *field pattern*. When evaluating the far field pattern for a particular antenna, it's customary for the antenna designer to provide information regarding the value of E in the θ direction (E_θ) and also that of E in the ϕ direction (E_ϕ). A field pattern graph would not, however, be terribly useful for an 802.11 network designer or support engineer. That's because the number of watts per square meter isn't a tremendously useful value when considering real-world, on-site issues. It would be like putting a speed limit sign on the highway reading "Do not exceed 200,000 Watts." Perhaps there would be a corresponding traffic law indicating that this speed limit only applied to vehicles of more than 13,000 Newtons. If only smart physicists were allowed to drive this might work, but this scheme just wouldn't be practical for the rest of us! Neither is it practical to use the raw measurement units of charge and energy to represent field strength outside the realm of study and research.

The field strength surrounding an antenna is presented relative to the field strength that would be present if a theoretical isotropic radiator were given the same power as the real antenna being considered. If the real antenna produces field strength in some particular direction that's twice as powerful as the isotropic radiator would then that direction is simply measured as being 3 dBi. The "i" in "dBi" stands for "relative to isotropic" and means that the dB ratio is relative to the theoretical value that would be obtained with an isotropic radiator. A graph of the dBi values is created, called a polar coordinate graph, and this is what an antenna vendor provides to specify the signal pattern for a particular model of antenna.

Polar Coordinate Graphs of Antennae Field Strength

A polar coordinate graph is a representation of the field strength surrounding an antenna relative to the field strength that would be present if the antenna were a theoretical isotropic radiator having equal field strength in all directions. The most typical representation is based on the relative strength of the electric field (as opposed to the magnetic field) surrounding the antenna and this type of representation is called an "E"-Graph (because "E" is the vector symbol for the electric field). An alternative representation is called the "H"-Graph which is based on the relative strength of the magnetic field (because "H" is the vector symbol for the magnetic field). "E"- and "H"-Graph types will be discussed later but you may be interested to know that the examples presented in the text are all "E"-Graphs as would typically be used in an antenna vendor's specifications.

There are two types of polar coordinate graphs used to show the field strength of an antenna. One represents the side-view of the space where the antenna is mounted and is called the Elevation Cut. The other represents a top-view, looking down on the space where the antenna is mounted and is called the Azimuth Cut.

To better understand these two views, let's first consider the Elevation Cut. Imagine you're standing in an empty warehouse as shown in the two pictures below in Figure 4.7. On the left you see an omni-directional antenna mounted to the ceiling. On the right a directional antenna has been mounted to one of the posts. These side-views of the space are representations of the Elevation Cut.

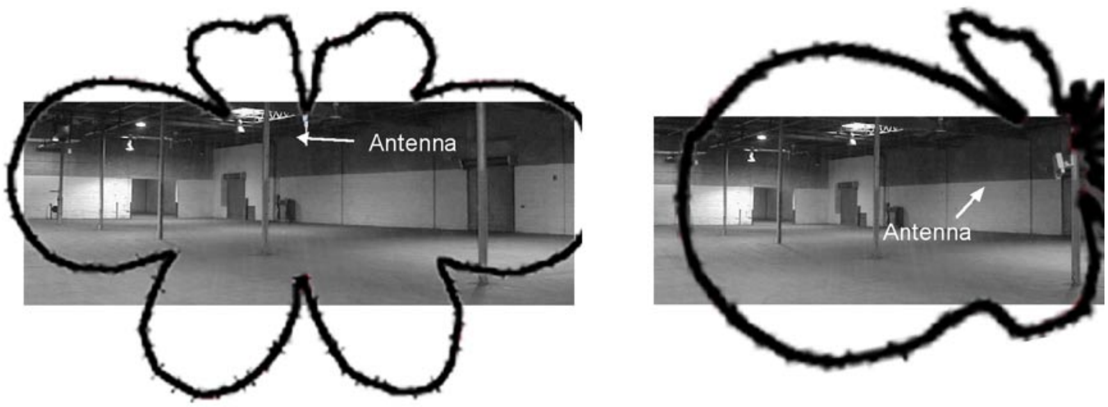


Figure 4.7 The Elevation Cut View of Antennae in a Warehouse

The pattern drawn on the picture represents the intensity of the signal strength for each antenna. The pattern represents a cross-section of the three-dimensional volume in space in which the field is propagating. This view is cut through the center of the three-dimensional volume and that's why it's called the Elevation Cut. The Elevation Cut view is, simply, a side-view of the field strength pattern.

If you were in a helicopter looking down on a building's floor plan, the cross-sectional view of the pattern from this perspective is called an Azimuth Cut. Figure 4.8 (below) is a floor plan for a building with the Azimuth Cut of a directional antenna drawn onto it. The directional antenna is mounted on the wall at the bottom of the picture. The Azimuth Cut is simply a top-view of the antenna field strength pattern.

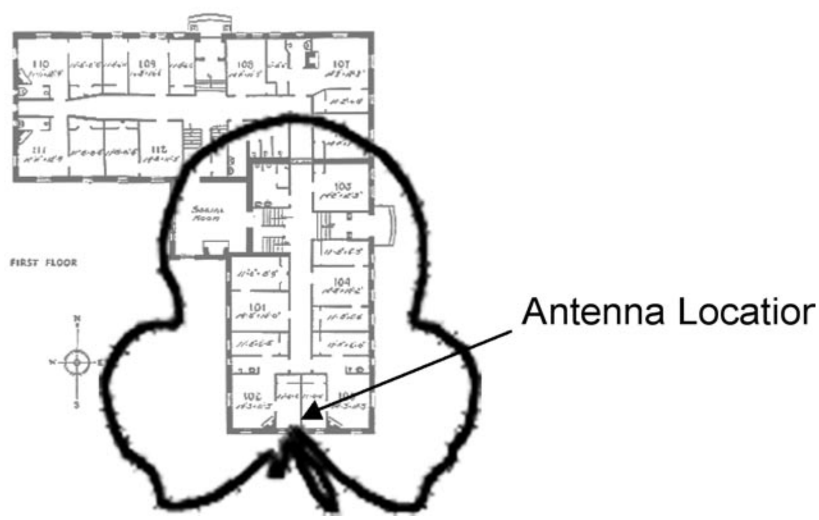


Figure 4.8 The Azimuth Cut View of a Directional Antenna

Antenna manufacturers provide signal strength graphs using polar coordinates to indicate the signal strength in different directions for their various antenna models. There will be two polar coordinate graphs for an antenna. One graph will show the signal strength as a side-view (called an elevation cut), and the other will show the field pattern as if you were looking down on it from above (called an azimuth cut). The following polar coordinate pattern graphs (Figure 4.9) are for a ceiling-mounted omni-directional antenna.

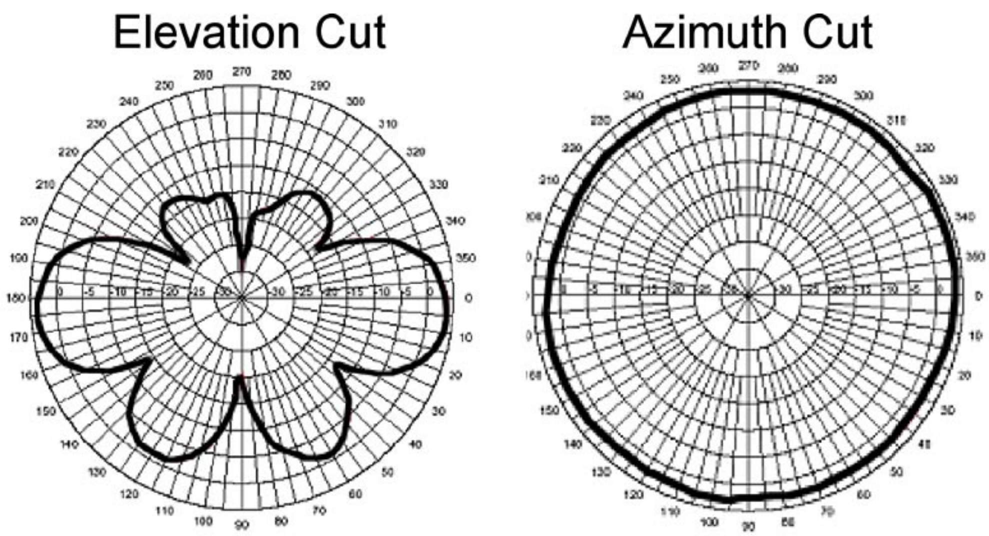


Figure 4.9 Polar Coordinate Graphs for an Omni-Directional Antenna

Notice that the Elevation Cut shows that the field is an odd shape when viewed from the side, but the Azimuth Cut shows that the field is circular when viewed from the top. By analogy, when a vertical cut is made through an apple (exposing the side-view; an Elevation Cut) it has a dimple in the top and bottom center. When an apple is cut horizontally (exposing a top-view; an Azimuth Cut) it essentially looks round.

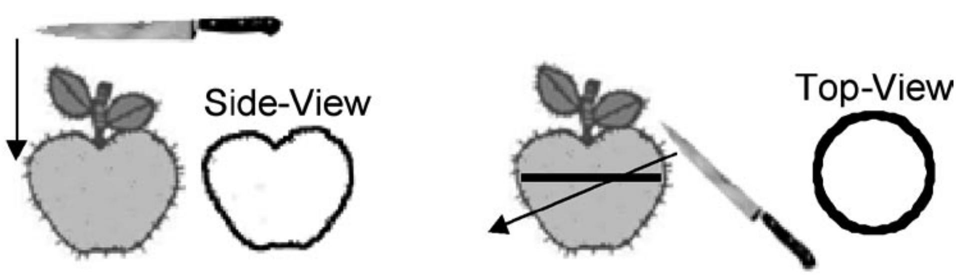


Figure 4.10 Vertical and Horizontal Cuts of an Apple

To more fully understand what's being shown by a polar coordinate graph, examine the values shown on the horizontal line through the middle of the graph. Figure 4.11 (below) is a close-up view of right-hand side of the Elevation Cut shown above. Remember that the dB values are relative to the theoretical isotropic radiator and are therefore referred to as "dBi".

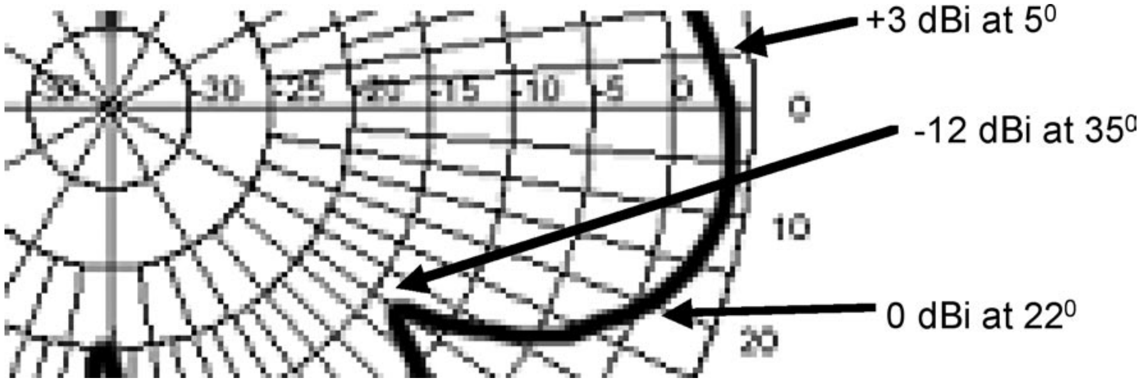


Figure 4.11 Close-up View of the Elevation Cut Polar Coordinate Graph

The polar coordinate graph is divided into a set of concentric circles, each one representing a dBi ratio. You see these circular gridlines labeled 0, -5, -10, -15, -20, -25 and -30 in the figure above. The radial lines represent degrees around the circle. In the case of an elevation cut the 0^O line is to your right, 180^O to your left, 270^O straight up, and 90^O straight down. In an azimuth cut the orientation of the coordinates remains the same but the 270^O direction represents the direction in which the antenna centerline is pointing. Of course, the direction is normally irrelevant for an omni directional antenna.

Notice, in the cross section (elevation cut) how the pattern has more signal strength in the downward direction (from the 0^O and 180^O centerline). The elevation cut represents the view you would have of the antenna's pattern if you were standing on the floor looking at the antenna from across the room. The elevation cut is the cross-sectional view as shown in Figure 4.12 below.

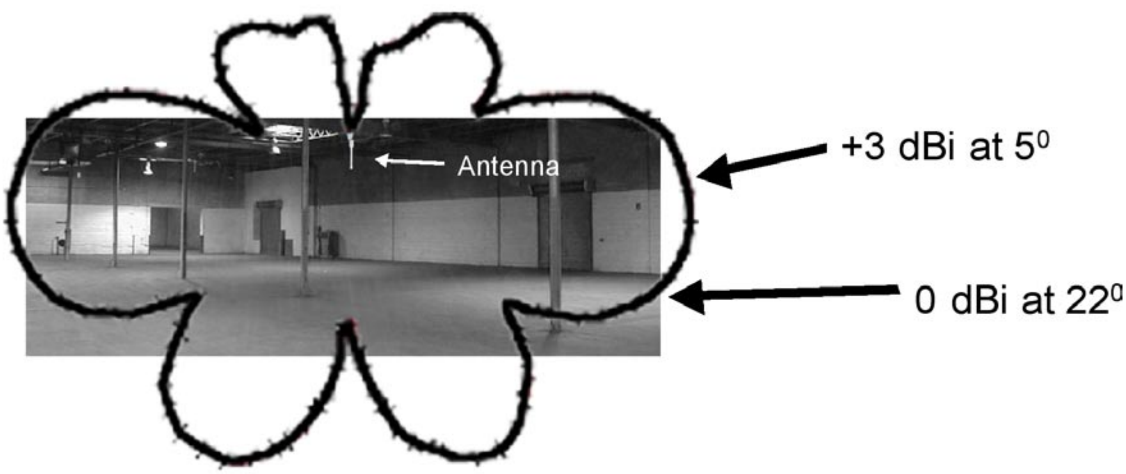


Figure 4.12 The Omni-Directional Elevation Cut Seen in the Warehouse

Figure 4.12 shows the view of the empty warehouse with an omni directional antenna mounted on the ceiling. The Elevation Cut graph represents the signal strength, relative to an isotropic radiator, that would be present in the warehouse from this orientation. Remember that you're looking at a cross-section of a three-dimensional thing. The three-dimensional field has the same cross sectional shape from any place you stand in the warehouse. Remember that the field pattern from the polar coordinate graph doesn't show you the extent of the field in terms of feet or meters but, rather, shows you what the signal strength would be relative to an isotropic radiator. The arrows in the diagram point to the same parts of the polar coordinate graph as in the earlier figure.

In the view-from-the-top (azimuth cut) you see the pattern is essentially round. The antenna has no "dead zones" and it transmits essentially equally in all directions. The azimuth cut is the top-view, looking down on the antenna from above.

The polar coordinate graphs for a directional antenna tell a different story for the cross-sectional view and view-from-the-top. Notice in Figure 4.13 (below) how the elevation cut seems to show that the directional antenna is pointing at the sky. That's only if you mount it in some silly way so that it is aimed straight up! You have to mentally adjust the information shown on a polar coordinate graph to fit with the reality of the antenna installation.

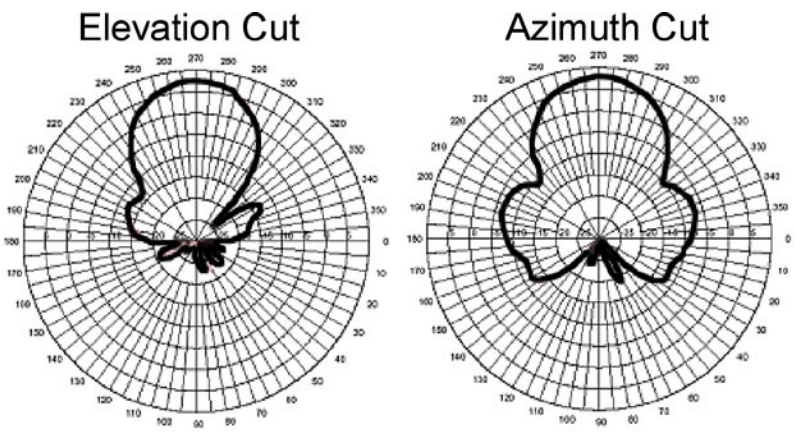


Figure 4.13 Polar Coordinate Graphs for a Directional Antenna

If you mentally rotate the elevation cut in the counter-clockwise direction (to the left), you'll find that the graph now shows what the pattern would be like if the antenna were mounted on a wall or a pole, pointing to the left, as shown in the rotated view of the elevation cut (Figure 4.14 below).

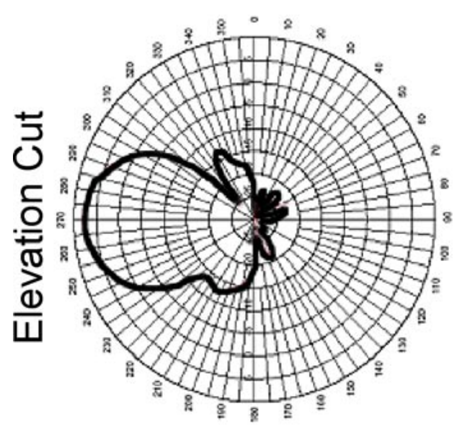


Figure 4.14 The Elevation Cut Rotated to the Left

If an antenna with this polar coordinate graph were mounted on a pole in a warehouse, one could imagine that the signal strength would be represented as in the picture below in Figure 4.15.

Figure 4.15 The Directional Antenna's Elevation Cut Seen in the Warehouse

The "E" Graph and the "H" Graph

Recall that the vector symbol for the electromotive field is "E" and the symbol for the magnetic field is "H". Recall, too, that these two aspects of the electromagnetic field operate at right-angles to each other. You will see polar coordinate graphs referred to as "E" Graphs or "H" Graphs, depending on whether they represent the electromotive force relative to an isotropic radiator, or the magnetic force relative to isotropic. Frankly, the "E" Graph is the most common, and most useful. But, if you encounter an antenna manufacturer or other reference using the "H" Graph it's a simple matter to mentally convert one format into the other. Remembering that the E and H vectors are 90^{O} apart, it turns out that an antenna's "E" Elevation Cut is the same as its "H" Azimuth Cut. In like manner, an antenna's "H" Elevation Cut is the same as its "E" Azimuth cut. You can visualize the reason for this equivalence by thinking about how your view of the field changes when you move from the Elevation view to the Azimuth view. These two views are 90^{O} apart, just as the "E" and "H" vectors are 90^{O} apart.

Half-Power Beam Width

The magnitudes of E_θ and E_ϕ in the far field vary inversely as the distance. At any point in space around an antenna there is a particular E_θ value that has decreased by 3 dB and is half as powerful as the maximum signal. This value is represented as E_θm. When a Poynting vector is drawn in the direction of E_θm, an angle is defined that's called the Half Power Beam Width. A side-view field pattern graph shows a representation of a cross-section of space on which the x and y axes lie and in which the antenna element is radiating outwards (towards positive and negative x). The plot shows the Half Power points. The angle between the Poynting vectors to the Half Power points define an angle called the Half Power Beam Width (HPBW), which is a standard measure of an antenna's directivity, as represented in Figure 4.16 below.

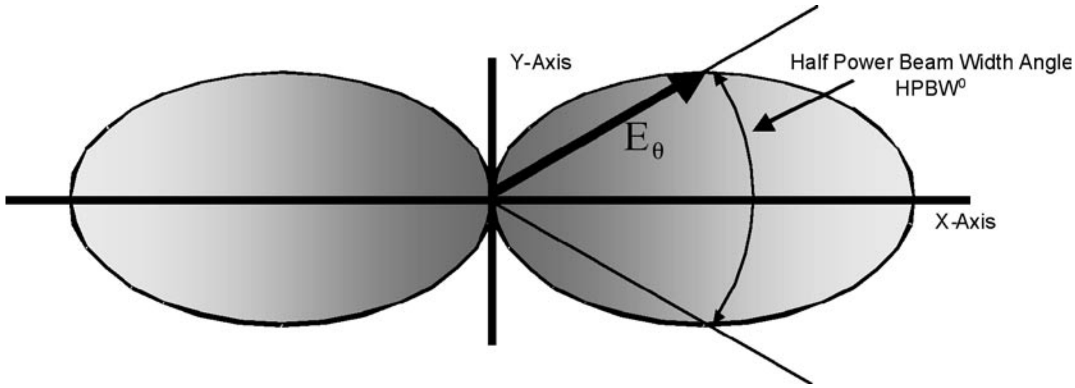


Figure 4.16 Antenna Field Pattern and Half Power Beam Width Measurement

Half-Power Beamwidth on a Polar Coordinate Graph

The Half-Power Beamwidth (HPBW) can be measured on a polar coordinate graph. There is a vertical HPBW (typically simply called the "vertical beamwidth") measured on the Elevation Cut and a horizontal HPBW (typically simply called the "horizontal beamwidth") measured on the Azimuth Cut. Since most indoor 802.11 applications confine their anticipated transmission areas to a single floor of a building the vertical beamwidth specification is sometimes ignored (although it should not be!). The horizontal beamwidth, taken from the Azimuth Cut, represents the field strength, relative to isotropic, as seen from a top view, which is just like looking at a floor plan.

Consider the Azimuth Cut shown below. Remember that this is a view-from-the-top. You first locate the point where the power is the greatest. In the example shown in Figure 4.17 (below), this is the point where the polar coordinate graph shows the maximum gain of 7 dB. Remember that a 3 dB reduction results in a 50% reduction in power. Studying the concentric circles on the graph you can see the 0 dB and 5 dB grids. Half-power is at 4 dB (because 7 dB minus 3 dB equals 4 dB). The points at which the field has decreased by 3 dB (hence, gone to half-power) define the Half-Power Beamwidth.

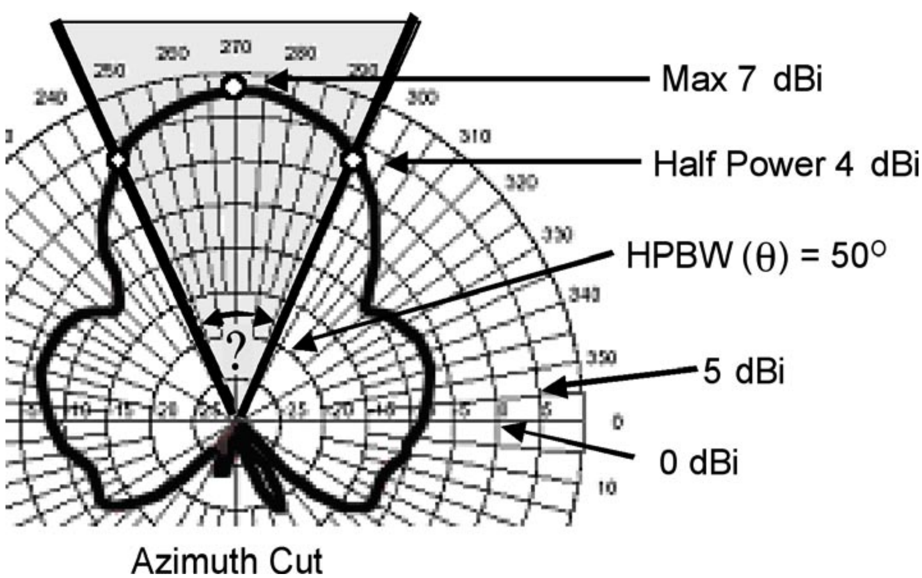


Figure 4.17 Identifying Half-Power Beamwidth (HPBW) Points

Every directional antenna will include a "beamwidth" specification, which is actually the Half-Power Beam Width (HPBW) angle. When talking in terms of beamwidth, it's common to describe a *horizontal beamwidth* and a *vertical beamwidth*. The horizontal beamwidth comes from the Azimuth Cut and the vertical beamwidth comes from the Elevation Cut. When a directional antenna is mounted on a wall or pole it's common to point it slightly down. This can be accomplished by physically positioning the antenna in its mounting bracket or the manufacturer can build in a *down angle*. The down angle is provided so an antenna can be mounted flat against a surface but still aim its beam slightly downward. The figure below (4.18) shows how the polar coordinate representation of the signal strength and the beamwidth would appear.

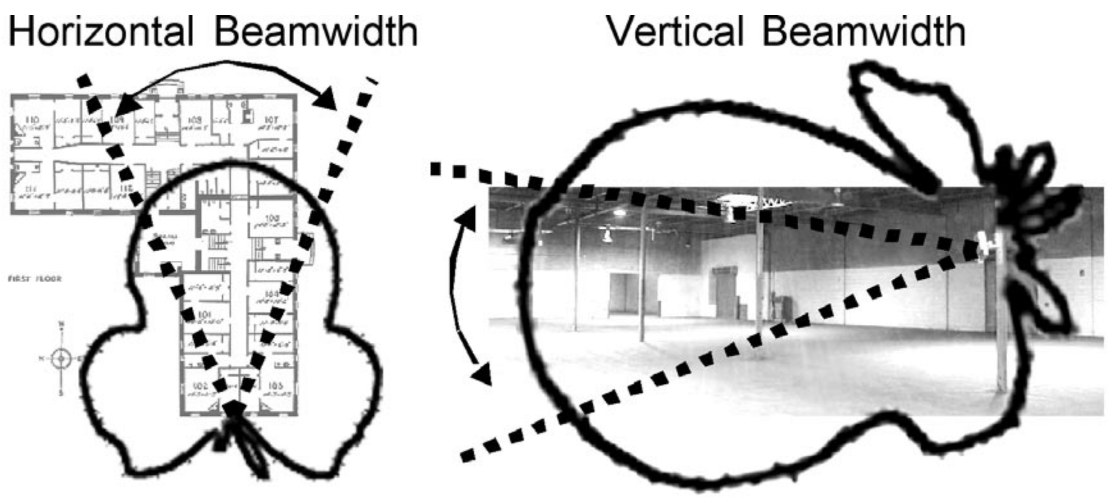


Figure 4.18 Horizontal and Vertical Beamwidth for a Directional Antenna

Full- and Half-Wavelength Antennae Beamwidth

Figure 4.19 (below) represents the field pattern for a 1-wavelength (λ) antenna. A 2.4 GHz 802.11 λ antenna would have a 12.5 cm radiating element, divided into two sub-elements making up the dipole structure. The signal would build up a pair of waves on the two dipole elements (as represented in the figure) giving rise to the "flattened" field pattern with a 47^O HPBW.

$$\lambda \quad E = \frac{\cos(\pi \cos \theta) + 1}{\sin \theta}$$

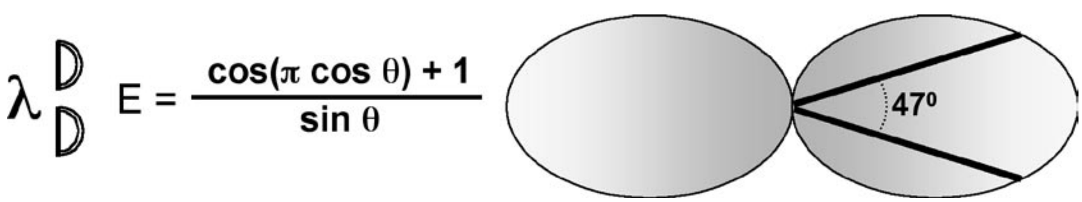


Figure 4.19 The Field Pattern for a Full Wavelength Dipole Antenna

Figure 4.20 (below) shows the situation for a ½ λ antenna. This would be the 6.5 cm "shortie" antenna sometimes seen on 802.11 PCMCIA NICs. Notice how the beam width has spread out to 78^O as a result of the way the wave is released from the two far ends of the dipole.

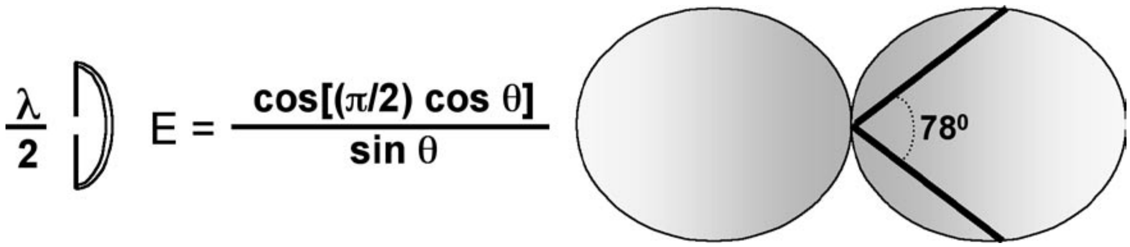


Figure 4.20 The Field Pattern for a Half-Wavelength Dipole Antenna

Now, if you take a moment to contemplate the two equations presented above (4.19 and 4.20) you'll see that the energy density (E, which is given in Watts per square meter, W/m^2) is calculated based only on the trigonometric functions sine and cosine (with a Pi thrown in because there is spherical propagation taking place). Unless you have some background with electromagnetism (or some other physics that uses lots of trigonometric functions) you might find this odd. "How," you might wonder, "can stuff that relates to triangles become units of W/m^2?"

Earlier, when we discussed the basic field equation, we talked about a "unit vector". An amazing thing happens in a physics equation when the length of a vector's arrow is set to the value "one".

Use of the Unit Vector

In the equations for electromagnetism we find the "unit vector" commonly used. This is because the "arrow" of a vector forms the hypotenuse of a right triangle relative to the axes of the coordinate system. Because the hypotenuse of the triangle in question is given as 1, the sine is equal to the length of the side opposite the angle and the cosine is equal to the length of the side adjacent to the angle. You can see this if you look at the basic trigonometric relationships. With the hypotenuse arbitrarily set to 1, the trigonometric relationships between the sine and cosine become representations of values on the x-, y-, or z-axis. When you see a reference to a sine ("SIN"), for example, you're seeing a representation of the y-axis value; cosine ("COS") represents the x-axis value. Consequently, when the unit vector is used as the hypotenuse, the SIN and COS values can be treated as if they represented a particular unit of measurement (like W/m2 (watts per square meter), E, H, or anything else).

802.11 Site Considerations Related to Beamwidth

As an 802.11 network designer, it's important to be aware of each antenna's beamwidth. It's also important to not get overly excited just because some part of the floor plan is outside the angle specified for the beamwidth. Remember that the beamwidth angle is always presented as the Half-Power Beamwidth (HPBW) and that means there will be signal coverage suitable for communication outside the specified beamwidth angle. It will simply be less than the signal at the limits of the HPBW.

An 802.11 radio lowers its data rate when it's not able to transmit at its maximum rate (11 Mb/sec for 802.11b and 54 Mbits/sec for 802.11a and g). The details of how bits are encoded in the 802.11 RF modulated signal is outside the scope of this discussion. Nonetheless, you should know that a single bit can be represented by an RF analog pattern that is of a longer or shorter duration. The shorter duration encoded bits result in a higher data transmission speed. The longer duration bits result in the data transmission speed being reduced. The transmitting radio determines whether or not it is capable of accurate transmission relative to noise and errors and adjusts its data transmission rate accordingly. This change in data rate is transparent to the user of the radio (except, of course, they may see their files being transferred more slowly). Most 802.11 radio designers use more than simple signal strength measurements to determine when to shift to lower data rates.

A Challenging Beamwidth Question

The Professor and the Chauffeur are discussing 802.11 networking. The Chauffeur has been learning, little by little, as he has had to sit through the Professor's lectures many times. The Chauffeur makes the following statement:

The ideal situation would be to have a user who is inside the HPBW of the access point antenna and the access point inside the HPBW of the user's antenna.

The Chauffeur illustrates his thought with the following diagram:

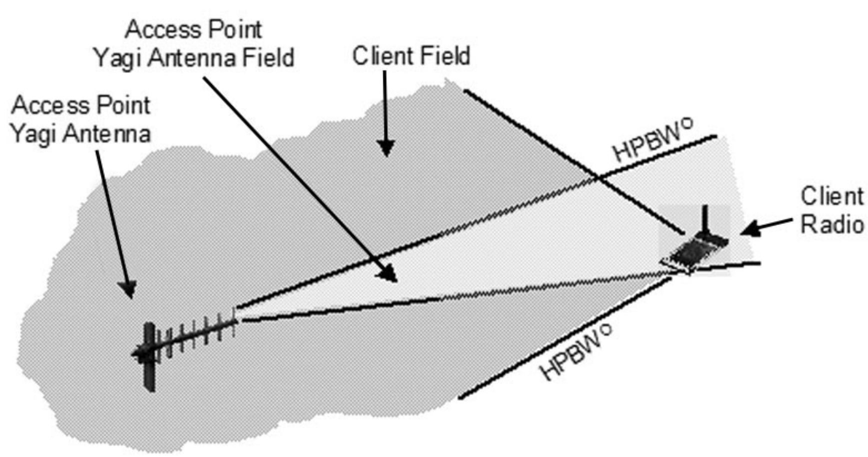


Figure 4.21 The Client and the Access Point Are Within Each Other's HPBW Zone

The Professor says, "Interesting idea, but I'm sorry to say that you're completely wrong." The reason the Chauffeur is wrong is because the reciprocity theorem says that "If I can hear you, you can hear me (for the same input power)." Hence, if the client can hear the AP then the AP can hear the client, or vice versa. Considering the location of 802.11 client machines relative to the HPBW of the access point antenna is appropriate. Confirming that the access point is within the HPBW of the client antenna is unnecessary. This assumes, of course, that the client and the access point have the same input power to their respective antennae.

Signal Strength and Reduced Data Rate

As a general rule of thumb, when an 802.11 radio is receiving at less than -70 dBm the range is being approached where noise and environmental considerations are probably forcing the data rate to drop. Remember that the power is going to drop in alignment with the Inverse Square Law.

Assume that you're 10 feet away from the transmitting antenna of a typical access point rated at 100 mW which is 20 dBm. You measure the signal strength along the perpendicular center-line of the antenna (at the point of strongest signal). You find that the signal strength at this location is -10 dBm (which is, by the way, not unrealistic). If you now move up or down in space, to the point where you meet the extension of the HPBW angle, you can expect the measure -13 dBm (half of what was measured at the center-line). By the time you get 100 feet away from the antenna you can expect your signal strength reading (at the center-line) to fall to less than -60 dBm. Now it's time to be concerned about being outside the HPBW angle since you're only 10 dB away from dropping below -70 dBm. Fortunately, the area inside the HPBW angle has grown as you got further away from the apex of the angle and you may find that you simply can't position a receiver outside the HPBW when you're 100 feet away from the antenna.

It's very important that you recognize that the preceding description assumed the measurement of -60 dBm at a distance of 100 feet. While this is a plausible value for some interior environments it is not to be construed to be a generally applicable rule-of-thumb or suggested guideline. As they say in automobile advertisements "Your mileage may vary."

We can use the basic trigonometric relationships to make some general comments about access point placement in 802.11 wireless networks. Consider a typical access point with a 1-wavelength antenna (roughly 12.5 cm). This dipole will have a HPBW of roughly 47º. Figure 4.22 (below) shows an access point mounted on a 9-foot high ceiling with a user sitting at their desk near the access point. The annotations on the right-hand side of the figure are explained following the figure.

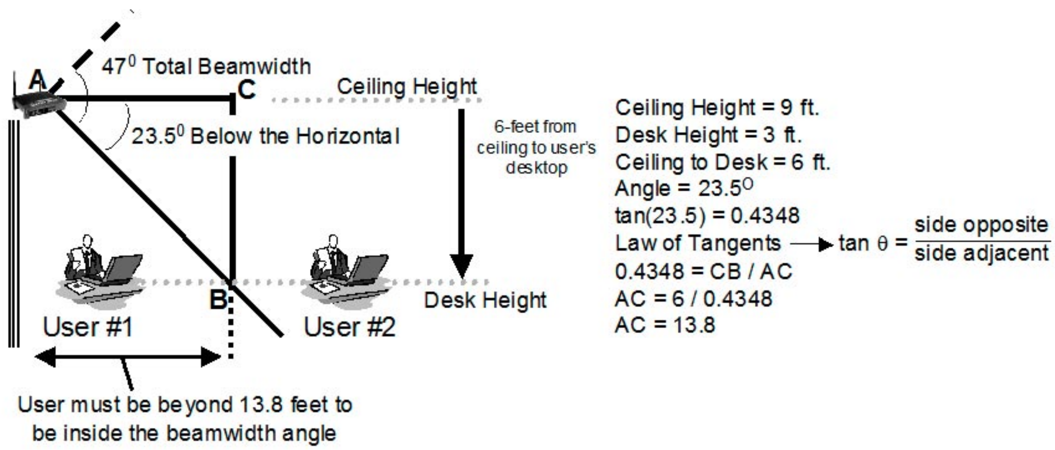


Figure 4.22 User #1 Is Outside the Beamwidth Angle of the Access Point

The ceiling in the figure (above) is 9 feet high. The top of the user's desk, where their notebook computer's antenna is located, is 3 feet high. The antenna, then, is 6 feet below the ceiling. Because the access point is mounted at the ceiling line (point A) the ceiling forms the centerline of the beamwidth angle. The bottom half of the beamwidth angle is 23.5 degrees (below the horizontal). The tangent of 23.5 degrees is 0.4348 (obtained using a scientific calculator). The tangent is the ratio of the side opposite the angle to the side adjacent the angle (Law of Tangents). In this case, that's the line CB (the 6 foot distance from ceiling to desktop) divided by the unknown distance from the wall, line AC. Rearranging the equation for the Law of Tangents to solve for the length of line AC yields: AC = CB / tangent or AC = 6 / 0.4348 and the answer (13.799) is rounded to 13.8 feet. User #1, therefore, is underneath the beamwidth angle and outside the 47O HPBW of the antenna. User #2 is beyond 13.8 feet from the wall and is inside the HPBW angle.

Physical Measurements Associated With the Polar Coordinate Graph

It's a common misconception that the shape drawn on a polar coordinate graph representing an antenna's field strength is somehow actually, and physically in some sense, a cross section of a three-dimensional volume in space. It is not. It's convenient to *casually* think of the three-dimensional shape conjured up from a polar coordinate graph as being some sort of volume growing out of an antenna because that makes for a generally workable mental model. Even earlier in this discussion the convenience of alluding to the model of the polar coordinate volume as some real thing beaming out into a warehouse was helpful in delivering the explanations. Now let's fine tune your understanding to see how actual, real-world signal strength (which can be thought of as field density in a volume in space) relates to the polar coordinate model.

The shape drawn on a polar coordinate pattern graph allows you to calculate the estimated field strength at some angle from an antenna. The graph shows dB relative to a theoretical isotropic radiator. For an isotropic radiator the signal propagating at right-angles to the antenna would be the same as that propagating at all other angles. If you were to measure the actual field strength at some distance from the right-angle plane relative to an upright antenna (the horizontal plane), then the polar coordinate graph could be used to estimate how that value would vary at some other angle. In the following figure (4.23) we see the elevation cut (side view) polar coordinate graph overlaid on a real antenna and a user. The polar coordinate graph is conceptual; it doesn't have a physical reality in the space around the antenna. It simply shows dB relative to isotropic at various angles.

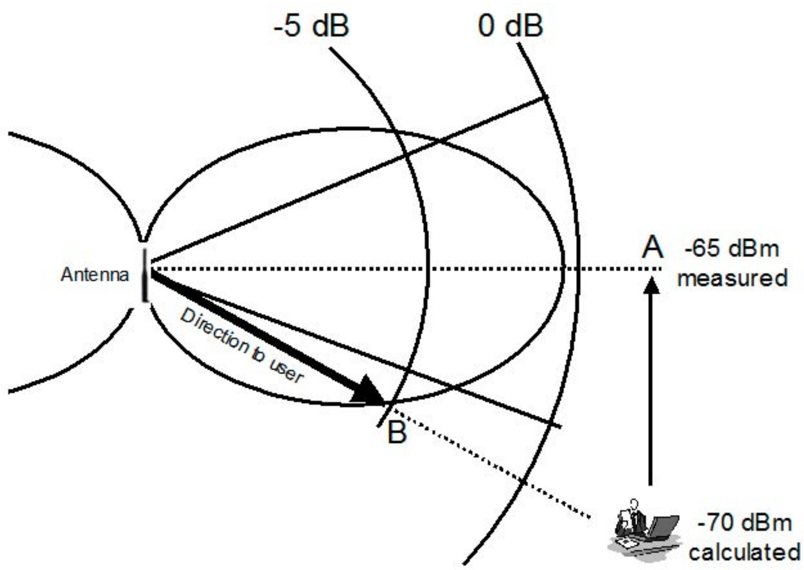


Figure 4.23 The Polar Elevation Cut as it Relates to a Real-World Situation

Assume that you measure the signal strength at some distance (point A) along the horizontal plane of the antenna. You determine that at that distance (perhaps 100 feet in a typical office environment) the measured signal strength is -65 dBm. You also determine direction from the antenna to the user and this is represented as an angle below the horizontal. If you overlay that angle on the polar coordinate graph, you can determine the point (point B) where the direction to the user intersects the field strength pattern. The coordinates on the example shown indicate that this point corresponds to the -5 dB attenuation level. You now know that any physical propagation of the field in that direction will be 5 dB less than along the horizontal plane. Since you measured -65 dBm at the lateral point corresponding to the user's position, you can calculate that the user would be receiving -70 dBm.

There's another physical reality to take into consideration in this situation that makes a real-world implementation of 802.11 perhaps not as immediately susceptible to poor performance in some locations as might be calculated. We'll discuss in detail later the fact that the RF wavefront doesn't conform to the free-space 3-dimensional shape and power distribution that would be calculated in theoretical terms. The environment is distorting the actual wavefront so that it's probable that the user will be in a zone that actually has more power than that contributed solely by the theoretical field pattern. It's also possible that the user could have less power available to them than calculated. Reflection, refraction, and diffraction of the RF signal is taking place due to the various metal objects, walls, furniture, and other objects in the environment. We'll discuss in detail how metal objects influence wave propagation. A typical 802.11 office environment may contain metal file cabinets, metal window blinds, and perhaps metal shelving units. Even the metal in desk or floor lamps influences electromagnetic wave propagation, and, of course, don't forget all of the metal in the computers in the environment too. Nonetheless, using the polar coordinate signal strength graph is a fundamental starting point in antenna positioning and RF coverage planning.

RF Modeling and Simulation

Advanced computer modeling and simulation software is available that applies the formulas associated with electromagnetic wave propagation to the floor plan of a building or to an outdoor area. Using simulation tools allows prediction of antenna coverage to within +/- 2 dB, a very accurate measurement!

Any time an 802.11 network is being designed it can be very valuable to perform a predictive RF simulation of the space and see what the electromagnetic wave propagation patterns would really look like.

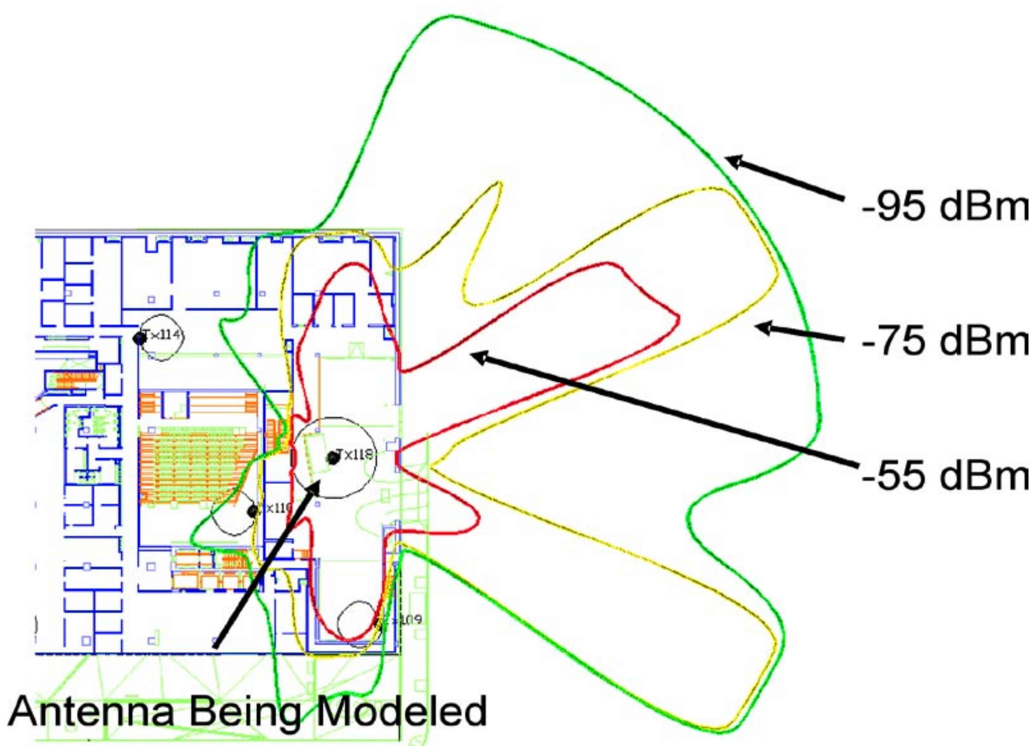


Figure 4.24 Results of an RF Simulation

The wavefront coming from the antenna being modeled is shown at the -55, -75, and -95 dBm propagation boundaries. You can see how the signal extends out from the exterior of the building with the strong -55 and -75 dBm lobes passing through the glass doors and windows. To the left of the antenna there is a wide metal stairway that is effectively blocking the signal.

Unlike an antenna's polar coordinate graph, the pattern drawn by modeling and simulation software actually represents an RF propagation boundary and shows us what the field would really look like if we could see it with our eyes.

Section 5: Electromagnetic Field Energy

A dramatic moment occurred in the history of physics in the 1860's when James Clark Maxwell combined the laws of electricity and magnetism and deduced the behavior of light. Light, of course, is simply a higher frequency form of the same "stuff" that makes up any electromagnetic transmission. Physicists call this stuff *photons* and although you may have associated photons only with light it turns out that they are considered a particle of field transmission in a category of particles called *bosons*. Bosons are referred to as "particles" but you must adjust your concept of what a "particle" might be when considering them. The typical concept of a particle is akin to what gets in your eye when you're in a dusty place. That is, the English language uses the term "particle" to mean something that has mass; it's an object, albeit a very tiny one. Physicists use the term *fermion* to describe a thing that's like that. Fermions are particles that have mass and they exhibit the quality that two fermions that have the same characteristics can't be close to each other. Electrons are fermions, as are protons. They are little tiny "things" that have mass, and they exist in the world in a form that can have a measurable diameter. Bosons, on the other hand, are referred to as *massless particles*. Where fermions are the class of particles that make up the "stuff" that we can see and touch, bosons are the "things" that convey force. The other three fundamental forces (gravity, the weak nuclear force, and the strong nuclear force) also exchange energy with particles having various esoteric names. Gravity is an attractive force between masses that is treated, by physicists, as being the exchange of a *graviton* particle. This doesn't mean that some little gravitons are swirling around in a soup, but in a way it does. Just like photons exchange forces within an electromagnetic field, gravitons (and other bosonic particles) do their work as if they were, perhaps, little tiny balls of something. The science of Quantum Mechanics has been trying to explain what that "something" is since the early 1900's and it continues to offer new, fantastic theories even today. There is a strong swell of thought in the physics community that multidimensional geometry, manifesting itself as if it were a cloud of vibrating string-like things, is at the core of everything. *String Theory* is considered by some to be the "Theory of Everything" and tremendous theoretical research is ongoing in physics departments at universities around the world.

The Particle Nature of the Electromagnetic Field

Bosonic particles (photons) are treated, mathematically, as if they were *point particles*. Photons have behaviors that resemble particles (they can be fired, one at a time, from a source, for example) but they also have behaviors that resemble waves (a stream of these "particles" will interfere with itself much like the wake from two boats would interfere on the surface of the water). Quantum Mechanics delves into the details of this dual nature and often refers to a set of photons as a "wave packet". Nonetheless, as per Einstein, the energy that they convey can be treated on the basis of its mass equivalent. It requires advanced mathematics to fully quantify the characteristics of photons, but the general ideas of mathematics revolve around the fact that equations can represent the effect of an infinite number of points over a particular square area or in a 3-dimensional volume. By understanding something about the characteristics of energy related to individual point particles (photons in the 802.11 field pattern) the concepts of the near field and far field can be explained, and those concepts are fundamental to understanding the physics of 802.11 wireless networking.

Field Power and the Inverse Square Law

Maxwell noted that the equations that had been formulated up to his time were mutually inconsistent. Previous research and experiments had led scientists to a partial description of the characteristics of the electromagnetic field but there were still missing pieces. Maxwell found that there was an aspect to the electromagnetic field that would decrease much more gradually than what had previously been thought. This led Maxwell to conclude that an electromagnetic influence could affect other charges that were very far away. He predicted the basic effects that we are familiar with today – radio transmission and, of course, 802.11 wireless LAN communication. It was known that certain aspects of the electromagnetic field decreased in inverse proportion to the square of the distance from the source. Maxwell found that another aspect of the electromagnetic field diminishes inversely in proportion to the distance, and not to the square of the distance.

It's common to hear about the "Inverse Square Law" that says "Signal power decreases in proportion to the inverse of the square of the distance." This means that if you measure the signal strength at some point and then move twice as far away (distance is doubled) the signal strength should be 1/4 as strong (power decreased by a factor of four). If you were three times further away, you would expect the power to be 1/9th as strong. This is a true statement but it refers to a characteristic of electromagnetic propagation that's the result of the spherical expansion of the propagating wavefront, and not due to some intrinsic characteristic of the electrical and magnetic field components.

It was Archimedes who first discovered that the surface area of a sphere was equal to the surface area of a cylinder of the same height and radius as the sphere, as seen in Figure 5.1 below.

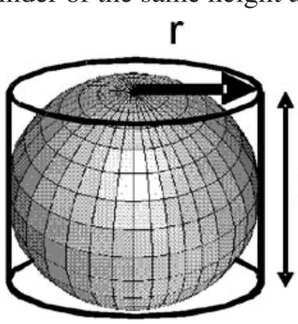


Figure 5.1 Determining the Surface Area of a Sphere

The ratio between the radius of a circle and its diameter is the constant π (Pi, roughly equal to 3.1415). Since the diameter of a circle is twice the radius, it follows that the circumference = 2 π r. The height of a cylinder that encloses a sphere will be equal to the diameter of the sphere which, therefore, is equal to 2r. Following Archimedes we can conclude that the surface area of the cylinder is the circumference of the circle multiplied by the height of the cylinder or $4 \pi r^2$. Watch what happens as the radius (r) doubles.

r = 1	surface area = 4π
r = 2	surface area = 16π
r = 3	surface area = 36π
r = 4	surface area = 64π

What you see is that if the radius increases by a factor of two, the surface increases by a factor of four. If the radius increases by a factor of three, the surface area increases by a factor of nine. Remember that the electromagnetic field is propagating outward and can be thought of in terms of a spherical wavefront. If the radius of the wavefront increases by a factor of two its surface area increases by a factor of four. The electromagnetic field is, therefore, spread out through an area that's four times bigger when the distance from the transmitter doubles. This is the Inverse Square Law.

Electric Field Strength Produced By An Individual Charge

We'll now consider the formula used to determine the electrical energy at a particular point. Since the total energy at the receiving antenna of an 802.11 device is the result of all of the individual points of energy, this fundamental equation becomes a key underlying component of the electromagnetic field itself.

$$E = \frac{-q}{4\pi\varepsilon_0}\left[\frac{e_{r'}}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{e_{r'}}{r'^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}e_{r'}\right]$$

Figure 5.2 The Strength of the Electric Field for an Individual Charge

This equation (Figure 5.2 above) is what a physicist might use to determine the strength of an electric field (E) produced by one individual charge. When the sum of all charges in a particular space are added together (vector addition) the result is the *charge density* in that space.

The first term ($-qe_{r'}/4\pi\varepsilon_0 r^2$) is, essentially, taken directly from a calculation called *Coulomb's law*. It tells us about a single electrical charge (-q; minus because it's a negative electrical charge). Charge is measured in units called Coulombs, where a single electron carries a charge of -1.602×10^{-19} Coulombs. The $e_{r'}$ term is a unit vector in the direction from the point where E is being measured. So, there's a charge, but the charge "feels like" a vector component of the charge because we're not "on" the charge itself but some particular direction (x-, y-, and z-axis location) away from the charge. This unit vector has a numerical value of 1 and simply points from the charge, back to the point, P, where the charge is being measured. The numerator of the first term, then, is the value of the charge and the direction in which the charge is located.

In the denominator you find, $4\pi\varepsilon_0 r^2$. The constant ε_0 (pronounced "epsilon naught") is the constant of permittivity of free space. Permittivity, as has been discussed, is the characteristic of free space or of a substance (such as water) that determines the speed of light in that medium. In a vacuum light (and any electromagnetic wave) will travel at 299,792,458 meters/sec. It will travel slower in dense medium which is the basis for signal refraction. The frequency and wavelength of transmitted energy are directly related to each other, and to the speed of light. We write "$v\lambda=c$" where v is the frequency of the signal, λ ("lambda") is the wavelength, and c is the speed of light in the medium. A 2.4 GHz carrier in an 802.11 network would have a wavelength of 12.5 cm in a vacuum, slightly less in the atmosphere (since c is slightly less). A 5.8 GHz carrier would have roughly a 5 cm wavelength in 802.11a.

The factor 4π speaks to the fact that the force is radiating outward from the charge in a spherical pattern and the surface area of a sphere is equal to $4\pi r^2$. Coulomb's law says, then, that if you have a particular charge being measured from a particular direction, the strength will be the result of the charge being spread out over the surface of a sphere of a particular radius and will be affected by the permittivity of the medium.

With regard to atmospheric propagation in free space (a vacuum) $\varepsilon_0 = 8.987552 \times 10^{-12}$ Farads / meter. The Farad (F) is the unit of capacitance. With 1 F of capacitance a single coulomb of charge can be held at one volt. A coulomb of charge is 6.25×10^{18} electron volts (the electrical potential of 6.25 billion billion electrons). The permittivity constant for the medium refers to the capacity of the medium to impede the propagation of a charge. If medium “A” can hold a greater charge than medium “B” the value of ε is greater for medium “B”. On a hot dry day, with a high pressure weather system over an area, the air is less dense than on a damp, foggy winter day when there’s a low front getting ready to bring rain into the area. These changes can have dramatic impacts on the propagation characteristics of an electromagnetic wave.

Under normal conditions the value of ε is very close to ε_0. and the denominator in Coulomb’s law, (given as $1/4\pi\varepsilon_0$) is treated as a constant (denoted by “k”) with a value of 8.987552 Nm/c^2 (Newton meters / speed of light in the medium, squared). As a result the first term of the energy equation could be rewritten as $-qe_{r'}/kr^2$.

Time Delay and the Retarded Wave

Maxwell discovered that Coulomb’s contribution was, as we’ve said, incomplete. This is where the next parts of the field energy calculation come into play. Let’s assume that a measurement is being done at some particular place that we’ll denote as (x,y,z). We’re measuring E at location (x,y,z) which, by the way, is located in the $e_{r'}$ (unit vector) direction from where we’re measuring. (There is no (x,y,z) in the formula, we’re just using this point to refer to where E is being measured). Now, the charge –q is assumed to be moving, vibrating in some particular way and at some particular frequency (resulting in some particular wavelength). An 802.11b or 802.11g device transmits a carrier frequency close to 2.4 GHz which is 2,400,000,000 vibrations (cycles) per second. This moving charge is producing a magnetic field and an accompanying electric field E.

The problem is that at (x,y,z) we cannot determine what –q is doing right now. The influence of –q on the space around it can not travel faster than the speed of light in the medium, denoted by “c”. Consequently, when we observe a particular –q at P we’re observing an effect that actually happened in the past. It didn’t happen very far in the past, but in the past nonetheless. This is a very mysterious thing and it’s referred to as a *retarded wave*. When a wave is measured, the measurement is always telling you how things were a moment ago. Of course, if we’re talking about electromagnetic measurements of energy coming from our sun then the time delay (called *retarded time*) is roughly 8-minutes. If we’re measuring microwave radiation coming from an object in a remote galaxy, then we could have retarded time of thousands of years (for an object that was thousands of light years away).

The fact that the wave is being measured with retarded time is accounted for in the second term of the energy equation, r’/c. The figure (5.2) is reproduced again below to help you avoid flipping pages as you read this paper.

$$E = \frac{-q}{4\pi\varepsilon_0}\left[\frac{e_{r'}}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{e_{r'}}{r'^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}e_{r'}\right]$$

Figure 5.2 (repeated) The Strength of the Electric Field for an Individual Charge

The wave retardation factor (time delay) is r’/c. What appears in the formula is the *apparent* distance that the charge is from (x,y,z) at the point of measurement.

If you think about a basketball player standing in one spot and dribbling a basketball you can get a picture of what's being represented here. The basketball is going up and down. It's "vibrating" and the charge –q also vibrates up and down in the radiating element of an antenna. Imagine that the inhabitants of the Doodle Galaxy, 1000 light years distant, have invented a fantastic telescope that lets them observe day-to-day activities here on earth. On some particular day, 100 years in the future, one of their astronomers is very excited to report that they have spotted an earthman demonstrating how to dribble a basketball. Of course, it took the light from Earth 1000 years to reach the Doodle Galaxy so the Doodle astronomers are seeing an event that took place 1000 years in their past. If they tried to perform calculations on the motion of the basketball they would be assessing an event that happened 1000 years earlier. In the meantime the basketball player would have grown old and died. The ratio r'/c adjusts the calculation to account for the fact that the –q charge was actually measured in retarded time. But this is simply a ratio and we're going to use this ratio to add an adjustment factor to Coulomb's result. The thing that's going to be adjusted is the second part of the equation.

The Derivative of the Energy With Respect To Time

Continuing our examination of the Electric Field Strength equation (Figure 5.2) we're going to consider the $e_{r'}/r'^2$ in the third term. Remember that $e_{r'}$ is the unit vector representation of the –q charge's position relative to the point of measurement and r' is the apparent distance that –q is from (x,y,z) (since –q may have moved since the signal originated and we're only seeing it in retarded time).

To understand what's happening here we need to understand that the term $e_{r'}/r'^2$ term is being differentiated with respect to time. The second derivative of the final term $e_{r'}$ is being taken.

The first derivative represents the *rate of change of the expanding field* and the second derivative represents the rate at which the rate of change is, itself, changing. Multiplying the rate of change of the field (the first derivative of $e_{r'}/r'^2$) by the adjustment required for retarded time compensates for the fact that we're not seeing the charge "now". The final term adds the product of $1/c^2$ and the second derivative of the unit vector $e_{r'}$.

Perhaps if you're unable to sleep some night and you're diligently working through this fundamental formula for determining the strength of an electric charge –q at location (x,y,z) you may encounter a challenge that is not yet totally solved by the physicists of today. You see, the charge, when viewed as a point particle, has an electric energy E that applies to itself. That is "What is the value of E when P is at the point of the charge itself?" The problem with this is that r'= 0 is not allowed as a denominator and that would be the case if you tried to plug values into the field formula when (x,y,z) is at the point charge itself.

Effective Radiated Power

Determining the energy of an electromagnetic field at some particular point involves some complicated math and introduces some surprises. We'll discover that there are two different regions of influence related to the radiated power and they are ultimately defined relative to the field equation.

To understand this we first consider some level of input power applied to the transmitting antenna. This power (a value of E and I across the resisting antenna element) is converted into an electromagnetic field that radiates outward from the antenna. The power output of the antenna is called the *Effective Radiated Power (ERP)* and is typically measured in milliwatts.

Consider the law of thermodynamics that states "Energy cannot be created or destroyed, only changed in form." Power goes in, power comes out. The total power, manifested by all of the charges in the radiating field, is equal to the input power, less any loss incurred in the conversion from input power to ERP.

You can think of this as if you had a garden hose and water was flowing out of it at a rate of 2 gallons per minute. You would still get 2 gallons per minute worth of water if you attached a lawn sprinkler to the hose and distributed the water over a 500 square foot area. If you put a nozzle on the hose to wash the windows on your house, you would still get 2 gallons per minute; it would just squirt out farther in one direction and, of course, there would be some loss due to the resistive characteristic of squeezing the flow through the nozzle.

The effective radiated power from an antenna can, also, be shaped and directed using high-gain and directional antennae. When we talk about antenna gain we see that ERP can remain the same, but when it's "squirted out" in something other than an isotropic sphere the effect is that the field, in a specific direction, has more power than the same field would have had in that direction in the isotropic situation.

The Near Field and the Far Field

Considering the field equation (from Figure 5.2 above) it can be seen that the aspects of the field calculated in first two terms falls off inversely with the square of the distance (r^2 in the denominator) The last term does not fall of with the distance but, rather, is differentiated with respect to time. This gives rise to two different parts of a radiating field. Very close to the source there are effects that decrease very quickly. As you get further away these effects become so small that they no longer cause measurable changes in the field. Further away the effects of the last term of the field equation are more significant. These two areas are called the *near field* and *the far field*. Sometimes these are also referred to as the Fresnel Zone (the near field) and the Fraunhofer Zone (the far field). It should be pointed out that we'll see the use of the term "Fresnel Zone" when we discuss interference patterns later in this paper. The two terms are the same, but their meaning is slightly different.

When the terms that vary inversely as the square of the distance are removed from the field energy equation the result is the equation for energy in the far field and is shown below in Figure 5.3. You can see how the energy equation from Figure 5.2 has been transformed into the far field version by the removal of the two middle terms and the consolidation of the common factors in the denominator.

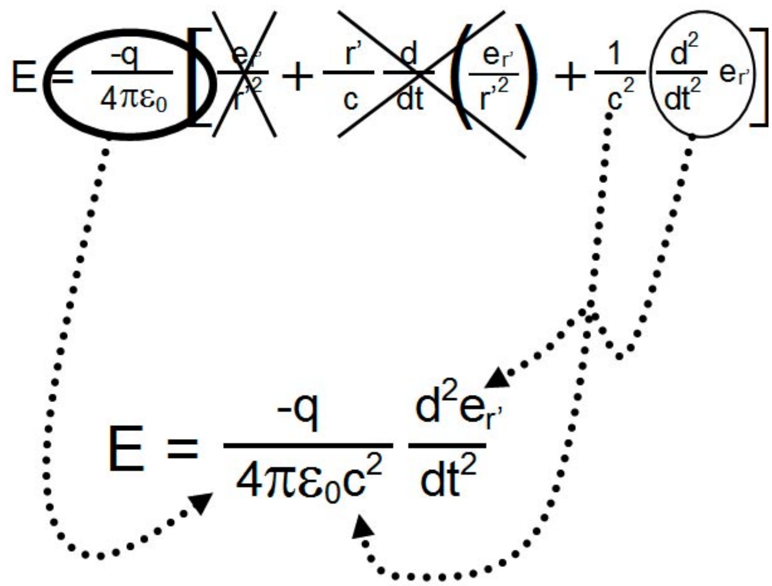


Figure 5.3 The Far Field Transformation of the Field Strength

The terms of the field strength equation that quickly become very small define characteristics of what's termed the *near field* or sometimes referred to as the *reactive near field*. The terms that are left after removing these quickly shrinking aspects define characteristics of what's termed the far field or sometimes referred to as the *propagating far field*. All of the effects that we're interested in with regard to practical assessment of an RF environment are in the far field and are embodied in the last term of the equation. An antenna designer, or a student of basic antenna design, will encounter the near field when considering the direct interaction of antennae in a coupled array or coupled antennae components to modify their combined transmission characteristics. *Coupling* refers to multiple radiating elements that are within each other's near field.

The classic *Inverse Square Law* states that "power varies inversely as the square of the distance". This "law" is a description of the expanding nature of the spherical wavefront, and relates to the density of the electric charge. If a sphere doubles in size its surface area increases by a factor of 4. The surface area of a sphere varies with the square of the radius. Therefore, if the distance from a field doubles the field density will decrease by a factor of four, varying inversely as the square of the distance. This is the essence of the classic Inverse Square Law.

Signal Acquisition from the Spherical Wavefront

There is a significant difference in the way a receiver is electrically and magnetically affected in the near field as opposed to the far field. One aspect of this difference comes from the fact that a spherically propagating wavefront presents out-of-phase portions of the wavefront to a receiving antenna. When the signal is received it's possible for the antenna to acquire portions of the original signal that are out of phase with each other and hence may decrease the receivable power. Diagram 5.4 (below) shows a transmitting antenna and for the purposes of this present discussion the wavefront will be considered spherical. In the real world the 3-dimensional signal space will be more "flattened" (even with a standard, typical 802.11 omnidirectional dipole antenna as would be found on an access point). Even with a "flattened" propagation area there is still a curved wavefront moving perpendicularly outward from the antenna. Notice, in the diagram (5.4) that the wavefront is made up of the points that lie at the same phase angle in the propagated signal. That is, there is a 3-dimensional surface, the “flattened” sphere, the surface of which is defined as the place where the phase angle of the transmitted signal is the same. You can see this in the diagram.

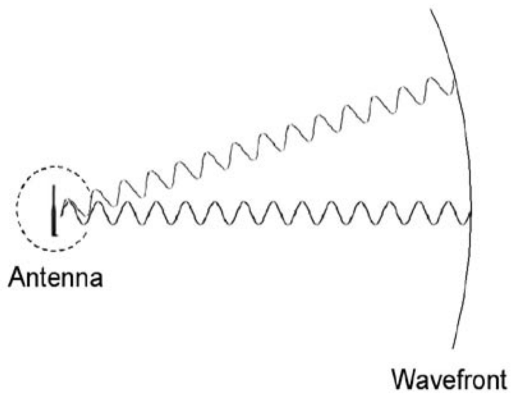


Figure 5.4 The Spherical Presentation of the Wavefront

In Figure 5.4 (above) you see the curve on the right representing the spherical wavefront. It’s moving to the right, away from the antenna. Notice that the sine wave in the diagram has just returned to zero from its negative half-cycle at the point where the diagram shows the wavefront. Since the wavefront is expanding in 3-dimensions, the inverse square law tells us that if the radius of the wavefront sphere were to double then the area on the sphere defined in degrees for some particular vertical and horizontal width would have four times the area. The overall surface area of a sphere is given by $4\pi r^2$ hence if r is doubled for any part of the sphere then the resulting surface area will increase by a factor of 4. It is this physical property of spherical surface area that is the basis for any inverse square relationships in the electromagnetic world.

Now let’s examine an implausible diagram. It’s implausible because it’s grossly disproportionate to anything that would be encountered in the real world. Nonetheless, we’ll start with this implausible scenario and then explain exactly what the implausibility is and how this impacts the realm of RF signal analysis. Figure 5.5 (below) shows a very long vertical antenna in the signal propagation area of the field.

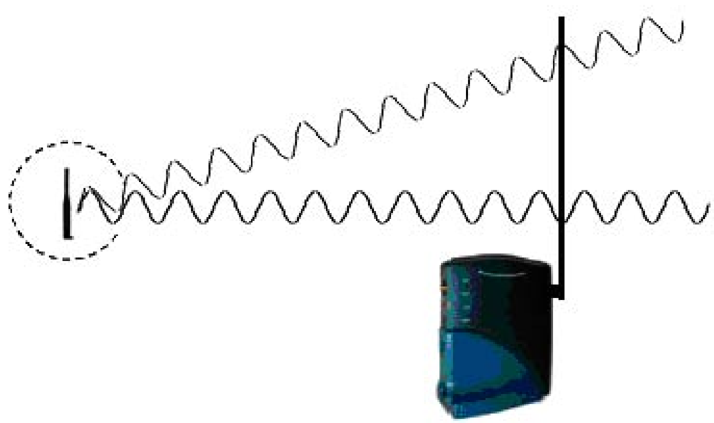


Figure 5.5 An Impossible Antenna of Unreasonable Length

When you study this diagram (Figure 5.5) carefully you'll see that the spherically expanding radiation pattern crosses the bottom of this particular antenna at the most negative part of the signal. For a sine wave this would be the 225º point. Notice that the radiated signal crosses the top of the antenna at the most positive part of the signal (the 45º point). The antenna is now receiving an equal, but oppositely charged input and the wave peak cancels the wave valley and the net signal to the receiving antenna is zero. This would be a very bad thing to have happen to your antenna since if you physically placed your receiving antenna at the point in the transmission path where the situation just described was happening, then you would not be able to communicate because the out-of-phase signals would cancel each other.

The reason that this is an impossible scenario is based first on the fact that we're talking about near field effects and you're probably not going to have a receiver within the near field of a transmitter in the 802.11 wireless LAN environment. This is because at 2.4 GHz or 5.8 GHz the near field is very, very small. How do we know that we're talking about near field effects? What we just described related to the spherical wave propagation ($4\pi r^2$) and that means we're talking about effects in the near field. We're going to see that the near field is so small that its significance is appreciated mostly by antenna designers and physicists, and not by engineers working with real-world equipment in real-world 802.11 wireless networks. In the far field, as we'll discuss shortly, the wave propagation is considered to act like a planar surface, and not like a spherical one.

The Boundary Between the Near Field and the Far Field

The boundary between the near field and the far field is at a distance R (in meters) from the source that can be approximated with the equation below.

$$R = \frac{2L^2}{\lambda}$$

In this equation L is the length of the antenna in meters and, of course, λ is the wavelength of the signal, also in meters. "R" (the boundary distance in meters) is called the Rayleigh distance. The equation is derived by considering a spherical propagation pattern that has expanded to a sphere that is sufficiently large that the wavefront that impinges on the tip of a receiving antenna is 22.5º (π/8 radians) out of phase with the wavefront in the middle of the antenna. This is considered to be a sufficiently "flat" surface of the spherical wavefront so as to consider it as if it were simply a planar surface. To get an idea of why this is true, consider Figure 5.6 below. In this figure you'll see the propagating signal considered at three separate angles: 0º, 8º, and 17º. These angles were chosen simply to make comparing and contrasting the wave phases clear. There is no special significance to these angles.

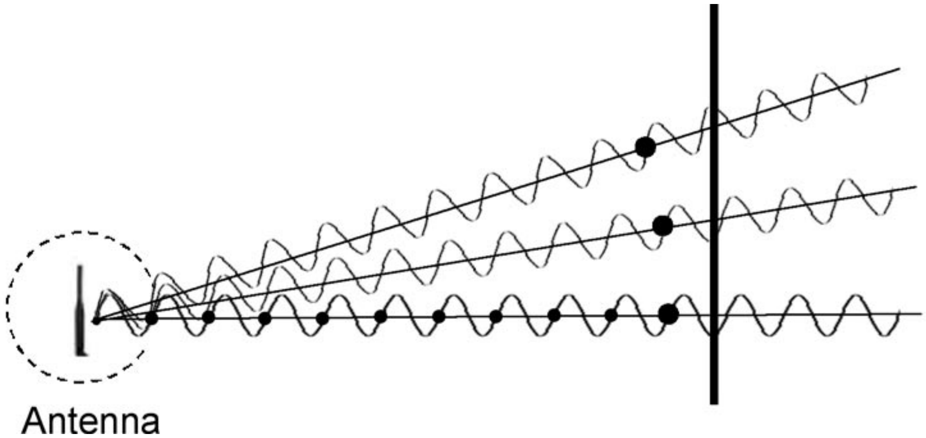


Figure 5.6 Out of Phase Signals Meeting a Vertical Antenna

On the 0º line you see that the start of each frequency cycle has been marked with a dot to serve as a point of reference. The large dot represents the start of the same cycle on all three paths. From this you can see, as we discussed earlier, that the vertical antenna intersects the spherical propagation field at different points for each angle.

One individual frequency cycle is measured in degrees, with a full cycle being equal to 360º. A sine wave peaks at 90º and hits its lowest value at 235º. A close examination of the three waves in Figure 5.6 is shown below in Figure 5.7.

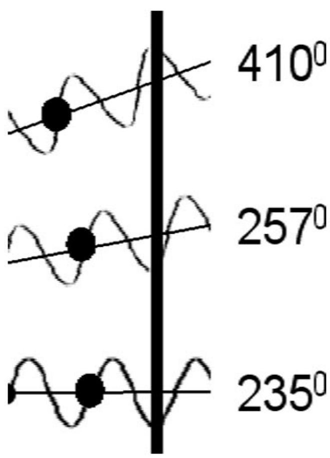


Figure 5.7 A Close View of the Out of Phase Waves

Notice how the 0º (bottom) wave hits the antenna at the lowest point in the cycle, the 235º point. The 8º (middle) wave hits at 257º, and the 17 º wave (top) actually hits the antenna in the next cycle. 410º is one full cycle (360º) plus another 50º. You can see that the antenna appears to be hitting the cycle just beyond the initial crest that cycles at 45º, hence it looks like 410º is a good approximation at the point.

An 802.11 access point typically uses a 1-wavelength antenna so L and λ are both equal to 12.5 cm. In this case $2L^2/\lambda$ = 25 cm (roughly 10 inches). You can see why we're interested in the far field for all practical purposes.

Characteristics of the Far Field

Some characteristics of the far field (which are, essentially, the opposite of those for the near field) include the following:

- In the far field the amplitude of the energy components attributable to the $d^2e_{r'}/dt^2$ portion of the field is stronger than the near field components that vary inversely as the square of the distance.
- The field strength varies inversely with the distance (1/r) and the influence of the inverse square field characteristics ($1/r^2$) can be ignored. Essentially, if you measure field strength and then move twice as far away from the source the field strength will go down by half.
- Field power is represented by wave amplitude and the complicated interactions between electrical and magnetic fields need not be taken into consideration in real-world measurements. Consequently, many of the influences caused by metal objects (like the shielding inside a computer) are minimized. Metal objects can still block or reflect the RF signal but the influences resulting from the oscillating RF energy creating a magnetic field in a metal object, and then that magnetic field having an effect on the surrounding field pattern or strength is negligible.
- Field strength is not very sensitive to changes in the location of the receiving antenna relative to the phase center of the transmitted signal. When the receiving antenna is moved through a distance of 12 cm (roughly 1 wavelength in the 2.4 GHz 802.11 band) there is negligible change in the field strength.
- Coupling and multiple reflections between adjacent antennae are not significant. The complicated relationships between phase differences in signals transmitted at the same frequency are how many directional antennae obtain their directionality, as is the case with the Yagi, Patch, or Panel antennae (to be discussed later in this paper). These, and other, interactions between transmitters are negligible in the far field.
- The wavefront of the expanding radiation pattern emanating from the source consists of transverse waves. A transverse wave field has a flat (planar) front with the wavefront being perpendicular to the direction of wave motion. This means that the spherical propagation characteristics that give rise to the inverse square relationship in the near field are no longer predominant. Rather, the wavefront seen by the receiving antenna is close enough to being planar that the spherical expansion characteristics can be ignored. There was a time when people thought the spherical Earth was flat. A receiving antenna in the far field has a similar perception regarding the radiating wavefront.

Considerations Concerning Near Field Interaction

While most practical real-world RF analysis takes place relative to a transmitter's far field, there are some thoughts to be considered regarding the near field. Remember that the boundary between the near and far fields is given as $2D/\lambda$ from the source which, as we've discussed, is roughly 25 cm in the 2.4 GHz 802.11 band. When you mount an 802.11 access point on an exterior wall directly on top of a concealed steel supporting beam, you're putting a big mass of metal in the near field. If the access point is closer than 10 inches to ceiling light fixture, again you've managed to get magnetic material into the near field. Finally, when you place your notebook computer on a metal file cabinet you get a magnetic substance into your PCMCIA wireless card's near field too.

When we discuss antenna coupling and re-radiation later on it will be explained how metal objects in an antenna's near field are energized by the signal power in a way that disrupts the radiated far field. Antenna designers use these disruptions in productive ways. Metal objects in your antenna's near field will probably result in degradation of transmission and reception capability.

The Reactive Near Field and the Radiating Near Field

The near field is actually thought of as having two parts: the *reactive near field* and the *radiating near field*. To understand the reactive near field we'll first consider the electrical and magnetic characteristics inside a simple piece of wire acting as an antenna. Actually, there are two pieces of wire in parallel, and very close together. An oscillating power source is connected between the two wires. We're talking about oscillations at 2.4 GHz or 5.8 GHz for 802.11 networking. At a particular instant the power source is pushing current into one wire, and pulling current from the other wire. An instant later the power flows in the other direction (as the waveform transitions from its positive half to its negative half). This means that electrons in the metal of the wire are being pushed and pulled inside the wire. The moving charges create a magnetic field and a signal radiates from the wire.
For one of the two wires we consider now the point at which the input signal polarity changes from "pushing" to "pulling". At this point the magnetic field around the wire collapses and re-introduces energy into the wire. So, not only is the energy in the wire being pushed and pulled, out and in, but so is the space immediately surrounding the wire in which the magnetic field is expanding and collapsing. Energy is being pulled not only out of the wire (during the negative phase) but out of the air surrounding the wire as well. This space, in which the energy is being pulled back into the wire during the negative signal phase, is the *reactive near field*. It extends a distance from the radiating element roughly equal to $\lambda/2\pi$ which, for 2.4 GHz 802.11, equates to 1.989 cm, a very tiny space. Nonetheless, when a notebook computer is sitting on a metal file cabinet the PCMCIA antenna can, in fact, be closer than 2 cm to the metal surface. The effects can be dramatic since the whole operation of the antenna depends on the correct manipulation of the electric and magnetic field. Metal objects (or other magnetically reactive objects) in the reactive near field should be considered extremely detrimental.

The radiating near field, extending from $\lambda/2\pi$ to $2D/\lambda$ (from 2 cm to 25 cm at 2.4 GHz) no longer returns significant energy back to the radiating element but the relationship between the electric and magnetic aspects of the signal is still strong. In this realm the spherical nature of the radiation pattern has not yet spread out sufficiently to act as a transverse wave and the unit vector directions for the field charges is predominant. This is unlike the far field where the field is considered to act like a transverse, parallel-path waveform. By designing the physical shape and size of the antenna itself with regard to these interactions in the near field, the characteristics of antenna gain and directionality can be implemented.

Antenna Gain and Directivity

Gain is the increase (or decrease) in signal strength measured relative to that which would be seen with a theoretical isotropic radiator and the unit is the dBi (decibels relative to isotropic). Physicists and antenna designers speak of *negative gain* but that's just another way of saying *loss*. Seeing a reference, for example, to "-3 dBi" one could say "a gain of minus 3 dB" or one could just as correctly say "a loss of 3 dB". Interestingly, one never speaks of "positive loss". These are simply quirks of language and have no real significance to the theories or facts themselves.

A directional antenna is built so that it doesn't transmit in all directions equally. Directivity (also referred to as *directionality*) is the characteristic of an antenna whereby the radiation pattern is focused in some direction or directions more than in other directions. The electromagnetic field strength is greater in some directions than in others. A parabolic dish antenna reflects all of its energy out in one direction, with none going out "behind" the antenna. This is the idea of directivity. In this realm we see that the idea of antenna directivity is accompanied by the idea of antenna *gain*. Gain is a relative measure of how a particular antenna's radiation pattern compares to the theoretical isotropic spherical pattern.

Of course, we can't really separate directivity and gain. In fact, the two terms, gain and directivity, are really alternate ways of thinking about the same characteristic of an antenna. If you want gain, you say "gain" and if you want directivity you say "directional" but there are only one set of wave properties in view.

We have considered the fact that when an electric current oscillates up and down in a wire an electromagnetic field is created that propagates outward in a direction perpendicular to the wire. From this characteristic the omni-directional dipole antenna is designed. The power delivered by the antenna is essentially concentrated near the generally circular cross-section of a circle on the horizontal plane cutting through the center of the antenna at 90^{O}. It is possible, through the design of the physical metal elements of the antenna, to cause the electromagnetic field to propagate more in one direction than in another. This results in two changes to the field. First, of course, there are areas into which the antenna transmits more power, and other areas that receive less power (as compared with either a dipole radiator or a theoretical isotropic radiator). Secondly, the areas into which the antenna is now radiating have more power than an omni-directional antenna would because, for the same input power, the total energy output of the radiating element remains the same (as compared to its dipole or isotropic counterpart) but now that energy is no longer spreading out in an omni-directional manner.

When we make general statements about antennae and RF measurements we often think in terms of a lossless state. For the purposes of discussion we consider the dipole radiator and purposely ignore the fact that if you put 100 mW of power into an antenna then slightly less than 100 mW will be transmitted out of the antenna. Some of that power is going to be lost in the process (to heat, and other physical factors). But, considering the lossless situation, you assume that of 100 mW goes in then 100 mW comes out. With that in mind you see that what comes out relative to the horizontal plane of a dipole antenna is greater than what comes out relative to the same plane for an isotropic radiator. Less energy comes out of the top and bottom of the dipole but the total 100 mW has to go somewhere. If you were to measure the power in the field pattern directly on the top of the antenna, you would find that it was less than to the sides. There is, therefore, a certain direction in which power is greater and, thus, even a simple dipole radiator is "directional" to some degree. Moreover, the power being radiated perpendicular to the radiating element results in a W/m^2 (watts per square meter) value that would be higher than a theoretical isotropic radiator. There is, therefore "gain" associated with a simple dipole radiator.

A vertical antenna is considered omni-directional because the electromagnetic field radiates horizontally outwards in all directions. The field does not, however, radiate in all directions vertically. An omni-directional antenna has signal gain compared to a theoretical isotropic radiator (with a truly spherical radiation pattern) as a result of the asymmetry in the radiation pattern. "Gain" is not the creation of additional power but, rather, a focusing of the radiation pattern from all directions into specific directions. Gain, therefore, implies that some degree of directivity will be given to the field. Technically, an omni-directional antenna does not radiate in all directions equally. As we've seen, the dipole radiator is not isotropic and the area above and below the antenna's longitudinal axes has a reduced field strength. Instead of being a spherical radiation pattern the dipole radiates in a pattern more like a torus (a toroidal pattern), as shown below in Figure 5.8.

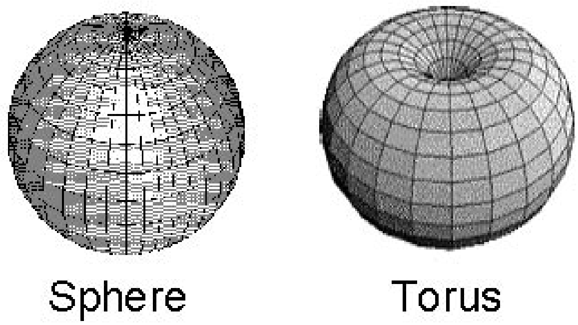


Figure 5.8 A Spherical Versus a Toroidal Radiation Pattern

It's important to understand what's being represented by the sphere and the torus above. This is not a picture of the electromagnetic radiation coming out of the antenna. Imagine that instead of an antenna we consider a vertical fluorescent light tube that someone might use in their garage or workshop, like the one pictured to the left. When you turn it on, the walls of the room are illuminated, and no light comes out of the top or the bottom of the tube. You do not see some kind of toroidal shape bulge out of the light fixture and expand as if it were a balloon filled with air. In the same way, you would not see a toroidal field around a vertical dipole antenna if your eyes could see in the 802.11 microwave wavelengths.

The sphere and the torus represent an imaginary surface in space. When you measure the power of the electromagnetic field at various points in space you find that the signal is weaker above and below the dipole antenna compared to what it would have theoretically been if you had measured at the same place for an isotropic radiator. The imaginary surface of the sphere is at a radius where some particular decrease in signal strength has occurred. Let's say, for the sake of discussion, that the surface of the isotropic sphere is the three-dimensional representation of the radius at which a 3 dB loss (half-power) has occurred. The toroidal shape, then, shows the same surface: the one at which a 3 dB loss would be measured. You can then recognize that the -3 dB radius is very short at the top and bottom of the antenna, since the toroid dimples in at the top and bottom. The shape of the three-dimensional representation shows relative signal strength, and not a fantastic picture of what a creature from Planet 10 would see in the microwave spectrum!

The consequence of the non-isotropic nature of the dipole antenna is that it has gain (increased power) relative to an isotropic radiator. We say "it has gain" and, of course, that's referring to the space volume perpendicular to the antenna. If we were speaking about the top and bottom of the antenna we would have to say "it has loss" (again, relative to isotropic). The same dichotomy is true for any antenna to which we attribute gain or directivity. When there's gain in one direction, or in one general plane, there must be equivalent loss in some other direction, directions, or planes. Directivity is simply a byproduct of directional gain. When an antenna is built such that it has significant gain in one direction and significant loss in the opposite direction, we say that it's a *directional antenna*.

There are many different ways for an antenna designer to construct the physical components of an antenna to achieve gain and directionality. Two fundamental concepts involve the use of a phase array or the construction of re-radiating elements (called *parasitic elements*). In the phase-array more than one radiating element is transmitting the same signal at the same time. They are, however, radiating signals that are not exactly in phase with each other. The result is that the signals cancel each other in some directions and constructively enhance each other in other directions. When parasitic elements are used there is only one element that is actually receiving an input signal from the transmitting circuitry. The other elements of the antenna are metal components that are in the reactive near field of the single transmitting element. They receive current from the excitation of the transmitting element's electromagnetic field and then they reradiate a field of their own. Because of the size of the re-radiating elements, and their placement relative to the transmitting element, they can be designed to both reflect the transmitted signal away from the undesirable direction and direct the transmitted signal towards the desired direction.

Phased Array Design Concepts

Consider two radiating elements placed exactly half of one wavelength ($\lambda/2$) apart. The figure below shows the top-view of two such radiators (as they would be seen if you were looking down at them from above). Assume that the exact same input signal is being fed into both radiating elements at the same time. This means that the same part of the transmitted waveform will be leaving each of the two antennae at the same time. Hence, when one antenna is transmitting the 0º point on the wave, so will the other, as shown in Figure 5.9 below.

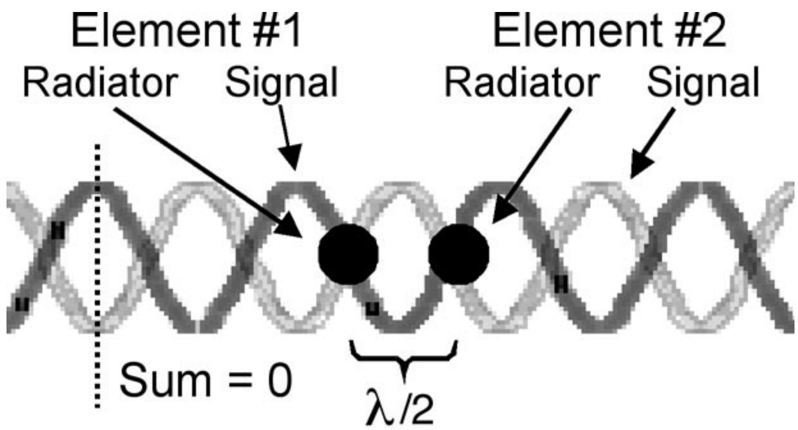


Figure 5.9 Top-View of Canceling Fields Parallel to the Two Radiators

Study the figure to see that Element #1, the left-hand radiator, at the moment shown is transmitting the start of a new wave (in the middle, half-way between the highest and lowest portion of the wave's amplitude). You can see that Element #2 is radiating the signal at exactly the same phase point. The same field is being generated by both radiating elements at the same time. Notice too, however, that the $\lambda/2$ spacing between the two elements results in the two transmitted signals being in the air exactly 180º out of phase. The maximum level for one signal appears at exactly the same time as the minimum level for the other. The sum of the two waveforms shown is zero: they cancel each other and no field energy is propagated to the left or right, along the line parallel to the two radiators.

On the other hand, consider the situation depicted below in Figure 5.10. These are the same two radiating elements, at exactly the same time as depicted in the previous figure, but this time the measurement is being made at a point perpendicular to the line between the two radiators. As seen in the figure below, in the perpendicular direction the two waves are in the air exactly in phase. The maximum and minimum points all line up. The sum of the two waveforms is twice the level of either waveform by itself.

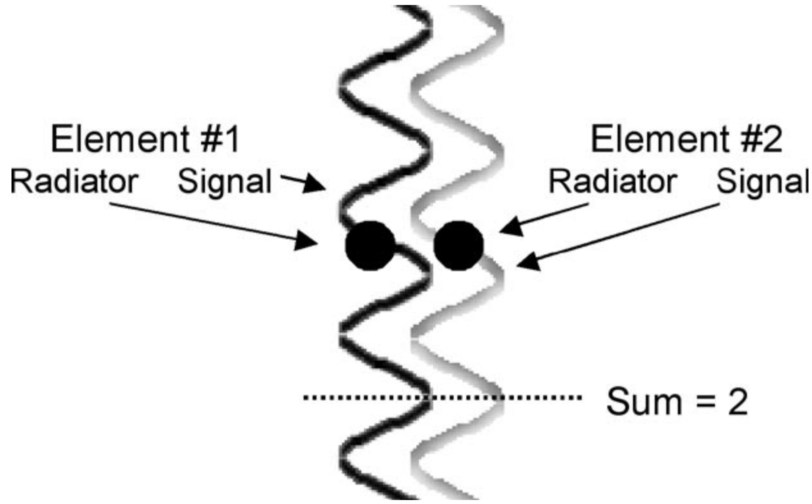


Figure 5.10 Top-View of Augmenting Fields Perpendicular to the Two Radiators

This is the principle of a phased array consisting of two radiators. The design has resulted in a directional beam with a signal gain of 3 dB.

You can see in Figure 5.10 above that the directional, augmented beam is shown going up, but it's also going down. In the direction that the antenna designer wishes the beam to point the beam is called a *front lobe*. The parts of the beam that go the opposite way are called *back lobes*. It is normally desirable to eliminate or minimize the back lobes and have the antenna focus its beam, as much as possible, in a single direction. Back lobes can be reduced by adding additional radiating elements at different distances and by adjusting the phase of the signal being transmitted from each additional radiating element. Moreover, the size of the additional radiating elements can be modified to affect the interaction of the fields. When additional radiating elements are introduced, the back lobe may be reduced but other lobes, called *side lobes*, come into existence.

Figure 5.11 (below) shows what the field pattern could look like for a phased array consisting of multiple individual radiating elements. There are a number of small side lobes but the back lobes have been effectively eliminated resulting in a highly directional beam.

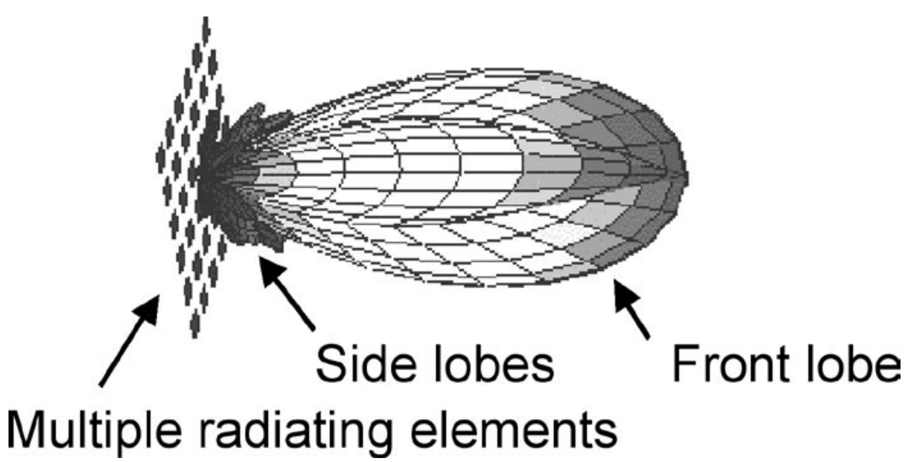


Figure 5.11 A Multiple Element Phased Array Field Pattern

Parasitic Element Design Concepts

The most commonly recognized antenna that uses the parasitic element design approach is the Yagi-Uda (or, simply "Yagi" pronounced "Ya-Gee" with a hard "G" as in the word "great"). The Yagi-Uda antenna (Figure 5.12 below) was named after the two Japanese engineers who originated the design. Recall that a metal rod in the reactive near field of a transmitting element will reradiate the electromagnetic field. By cleverly aligning the metal elements of the antenna, the re-radiating elements can serve to either reflect or direct the beam. This is how a Yagi antenna operates.

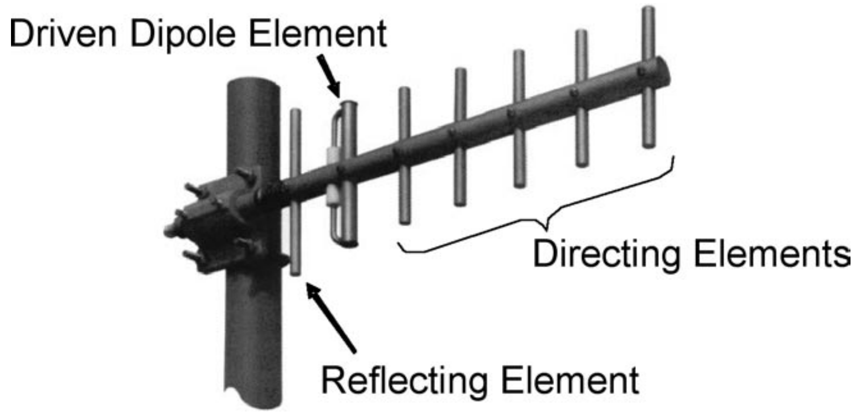


Figure 5.12 The Yagi-Uda Antenna

The driven element is typically a half-wave ($\lambda/2$) dipole radiator, fed with input signal at its center. The parasitic elements are not directly excited with input signal but, rather, reradiate energy received from the driven element. The reflecting element is typically about 4% longer than the driven element and directing elements are typically about 4% shorter. To achieve either the reflecting or directing effect the antenna designer selects the correct length and spacing for the elements.

Antenna Beamwidth and the Law of Reciprocity

Even a simple dipole antenna doesn't radiate equally in all directions. The measurement of the directivity of an antenna is called its *beamwidth*. Since we're dealing with three dimensions in the transmission volume there's going to be one beamwidth given for the view from the side (as shown in Figure 5.13 below) and another beamwidth given for the view from the top.

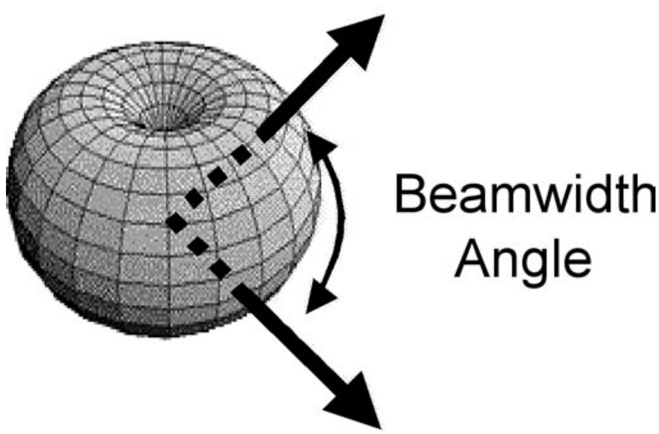


Figure 5.13 The Depiction of an Antenna's Beamwidth

An important fact to remember is that the beamwidth angle defines not only how the transmitted signal power will be dispersed but, also, how well the antenna will be able to receive signals. This is because, as provided in the Law of Reciprocity, the same characteristics that make the antenna directional (or provide it with gain) for transmitting, work exactly in reverse with regard to reception.

Section 6: The Huygens-Fresnel Principle

Christian Huygens wrote a treatise in 1678 on the wave theory of light. The ideas that he put forth were rooted in some very deep mathematics, well beyond the scope of the present discussion. It is important, however, to understand what Huygens postulated since it has been since proven to be a valid representation of a tremendously complex physical reality. Does electromagnetic wave propagation really work exactly as Huygens explained? No, electromagnetic wave propagation obeys quantum mechanical rules that allow superposed possibilities with only probability amplitude to describe whether or not some particular "reality" will be the one that's observed. Does electromagnetic wave propagation work in accordance with Huygens' explanations? Yes, in classical physics Huygens' equations correctly describe certain aspects of the behavior of a propagating field. Unfortunately, Huygens' principles have some shortcomings which were corrected later by Augustine Fresnel and then applied to Maxwell's equations by Gustav Kirchoff. Understanding Huygens' wave theory for electromagnetic field propagation provides the basis for not only formulating a better mental picture of what's actually going on in the air in an 802.11 environment but for understanding signal obstruction in accordance with an aspect of field diffraction known as the Fresnel Zone, which will be discussed.

Huygens wrote in the language of mathematics and appreciating his equations with any clarity requires the use of advanced calculus. Nonetheless, a model of what he described will be sufficient to understand his principles. Consider a source of an electromagnetic field (like an antenna) as shown below. For an isotropic radiator the field propagates outwards in a spherical manner and the figure (6.1 below) shows the spherical wavefront at some moment in time as it's moving outwards from the source.

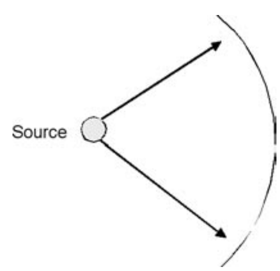


Figure 6.1 A Spherical Wavefront from an Isotropic Radiator

Huygens said that the wavefront itself could be considered to be a set of new source points, each one being the center of a new radiating field. For example, the figure below (6.2) shows one of the new source points (A) on the current wavefront and that point has generated a new spherical wavefront (B), called a *wavelet*.

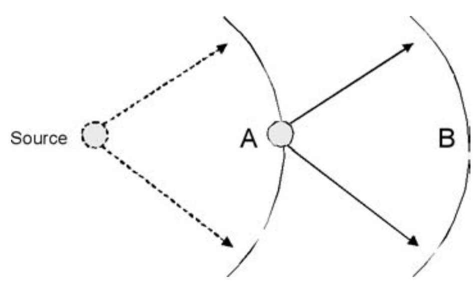
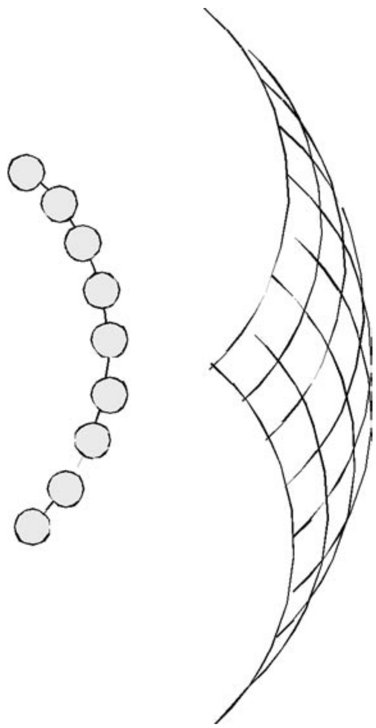


Figure 6.2 Each New Point Source Generates a Wavelet



Calculus allows the physicist to evaluate the infinite series and show that the wavefront of a propagating wave at any instant conforms to the envelope of spherical wavelets propagating from every point on the wavefront at the prior instant. The propagating wave front, then, can be conceptualized as shown on the left.

The first wavefront (far left) is considered as an infinite series of new field sources. Each one has its own spherical wave propagating outwards. The resulting field is represented by the interaction between all of the individual fields thus created, as seen on the right-hand side of the figure.

The problem with Huygens' original thesis was that while his ideas worked to calculate characteristics related to reflection and refraction but they don't account for why the wave propagates outwards. If each point generates a spherical wave then why does the wave tend to move outwards? Why doesn't it just sit there and jiggle, or some such thing? Another problem is that Huygens' principles don't account for diffraction, the bending of a wave around an obstacle.

Augustine Fresnel expanded on Huygens' original theory of electromagnetic wave diffraction (no pun intended). The resulting description of wave behavior is called the Huygens-Fresnel Principle.

Some physicists downplay the significance of the Huygens-Fresnel principle in favor of the newer theories of Quantum Electrodynamics (QED) but it is nonetheless true that the principle was born of truly inspired insight and is, without doubt, a forerunner of the newer discipline of QED. It could be said that the Huygens-Fresnel principle gives the right answers for the wrong reasons.

Applying the Huygens-Fresnel Principle in the 802.11 Environment

The integration of the sum of the wavelets produced by the points on a wavefront can provide a quantification of the characteristics of the expanding electromagnetic field. With this model in mind, consider what would happen to the wavelet integration if part of the expanding field were obstructed by, perhaps, a steel girder in a building or a metal file cabinet (or a wall, or anything else that is not transparent to the 802.11 signal). In fact, when an obstruction is present, the overall affect is that the shape of the field "bends" as depicted in Figure 6.3 below.

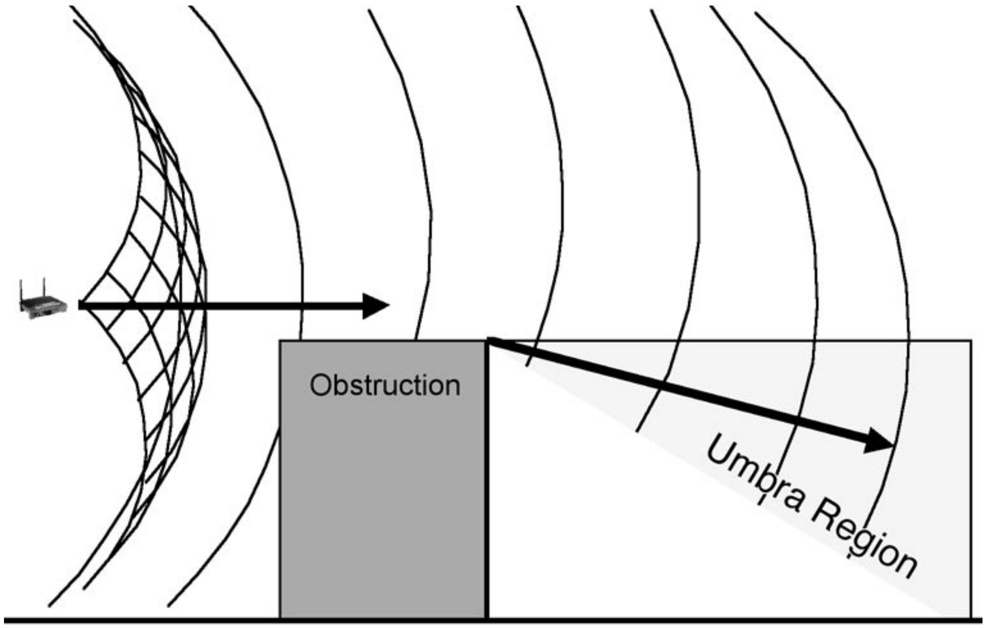


Figure 6.3 An Obstruction Causes the Wavefront to Bend

In fact, there is a sense (based on Huygens' idea of each point on the wavefront being a new source) that the obstruction is acting like a new antenna, bending the signal as if it were the source, as indicated by the line in the figure above (6.3). The expanding wavefront to the right of the obstruction is shown with its center on the obstruction's edge and no longer on the original antenna.

So, the Huygens-Fresnel principle tells us that an obstacle in the 802.11 transmission path will cause the field to diffract (bend) around it, analogous to the way a water wave creates eddy currents around the piling of a pier. Diffraction can be either a good thing, or a bad thing. It's good from the standpoint that it helps get RF coverage in an area; the field bends slightly around obstacles. It's bad from the standpoint that the diffracted field interferes with the original field and causes potential areas of reduced coverage. The shadow zone behind the obstruction is not in line with the top of the obstruction, but takes on a cone-like shape around the obstruction as the waves bend around the object. The *umbra region*, also called the *diffraction zone*, between the shadow zone and the cone of silence is a region of weak (but not zero) signal strength. In fact, the signal strength in the cone of silence typically doesn't reach zero since a certain amount of reflected signal is scattered from other sources causing the cone of silence to actually manifest some signal power.

Diffraction of the Expanding Wavefront

Diffraction is the bending in the direction of travel of part of a wavefront resulting from the presence of a radio-opaque object in the transmission path. This is a characteristic of electromagnetic radiation that can not be explained solely on the basis of a particle view of signal propagation. Diffraction is a phenomenon unique to the wave-like nature of the electromagnetic field. We leave it to research in quantum mechanics to suggest reasons why diffraction is consistent with both the particle and wave view of electromagnetism and, from the perspective of 802.11 wireless networks, think of diffraction only from the wavefront perspective.

Diffraction effects are most noticeable when objects are close to one wavelength in size (roughly 12 cm for an 802.11 2.4 GHz transmitter). By the time an object gets to roughly 60 wavelengths in size (7.5 meters) diffraction becomes negligible. This means that typical objects in any building's interior space will introduce diffraction. A tall cubicle wall, high shelving unit, or free-standing kiosk in an auditorium will all fall under the 7.5 meter height guideline. Horizontally, wall segments (not separated by doors) will typically be narrower than 7.5 meters and will act like horizontal barriers, also introducing diffraction at their edges.

In essence, waves propagate around electrically small (close to 1-wavelength) objects without any crisp shadows. The umbra region has sufficient energy to often go unnoticed by the casual user of 802.11 equipment.

The bending effect observed when a wave is diffracted around an object is the result of interference between different parts of the electromagnetic field. This concept is very mysterious and a concise explanation of the field behavior can only be expressed through some very complicated equations. The effect can be understood, however, by considering the basic characteristics of field propagation as expressed in the Huygens-Fresnel principle.

How Interference Relates To Diffraction

The details of the mathematics underlying Huygens-Fresnel diffraction are beyond the scope of this paper, and beyond the scope of the reader without a background in advanced math. As such, you're going to be faced with a mystery that you'll be able to fully appreciate from your reading, but it will remain mysterious nonetheless. We start by remembering that the spherical wavefront expands outwards as a series of Huygens' wavelets. The wavefront at some particular future time (t_2) will be the result of the interaction between all the wavelets produced by the wavefront at a previous time (t_1).

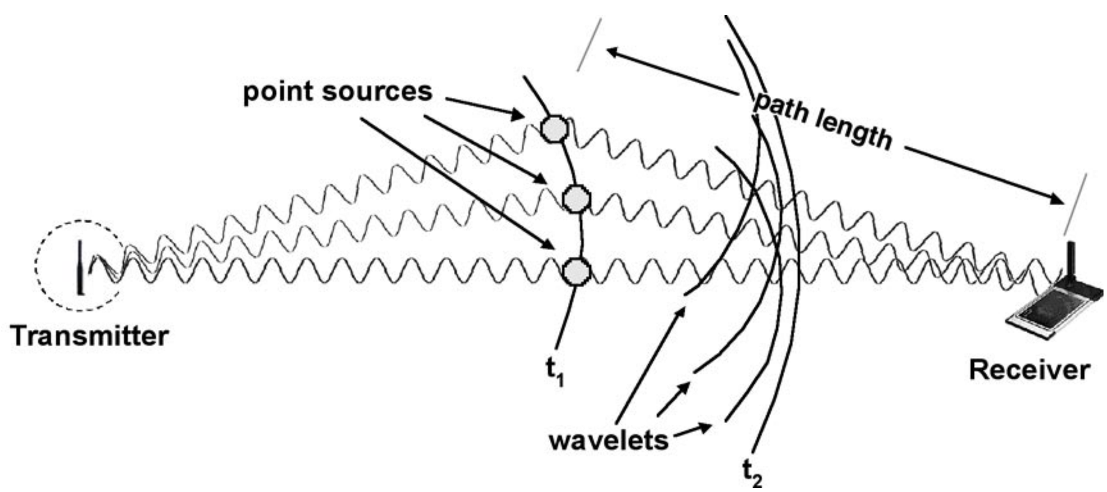


Figure 6.4 Wavelets Combining Out of Phase at the Receiver

There are three point sources shown at time t_1 in the figure above and we'll call them "top" "middle" and "bottom". Each point source produces its own spherically expanding wavelet and those wavelets combine to produce the new wavefront at t_2. In the overall picture the wavefront itself is expanding spherically with the transmitter as its center. The length of the radius (from transmitter to any point source on the wavefront at time t_1) is always equal and the effects of interaction between the wavelets always produce a new, spherical wavefront at time t_2. The view from the receiver's perspective is quite different. Notice in the figure that the path length from the top point source is shorter than the path length from the receiver to the middle or bottom point source. At t_1 all points on the expanding wavefront are in phase. From the view of the receiver the phase of the wavelet from the top point source has fallen behind the phase of the wavelet from the bottom point source because of the longer distance traveled.

Now if you contemplate the description just offered you should be thinking to yourself "Wait a minute! The receiver isn't going to see wavelets from each t_1 point source separately. The combined wavelets from t_1 will produce a spherically expanding wavefront that will pass across the antenna of the receiver and all parts of that wavefront will be in phase." Previously in this discussion we explained how the spherical wavefront was seen as a flat, planar wavefront as long as the phase is less than the Rayleigh phase angle of 22.5º. The planar presentation was an illusion (like the "flat Earth") due to the size of the receiving antenna compared to the curvature of the spherical wavefront. Here's the key to the mystery: The expanding wavefront does retain its spherical (and planar) characteristics, and Huygens' wavelets are a moot, theoretical aspect of propagation, until an obstruction blocks some of the point sources at t_1. When some of the t_1 wavelets can't propagate to t_2 then the differing path lengths are not just theoretical (from the view of the receiver), they are descriptive of the reality in the electromagnetic field.

When the path length taken by one component is ½ wavelength longer than another then the phase of the two components is 1800 out of phase and they cancel each other out. As the path length becomes longer it eventually reaches a point where an entire additional wavelength has been added. At this point the signal is back in phase with the original, albeit one wavelength further along.

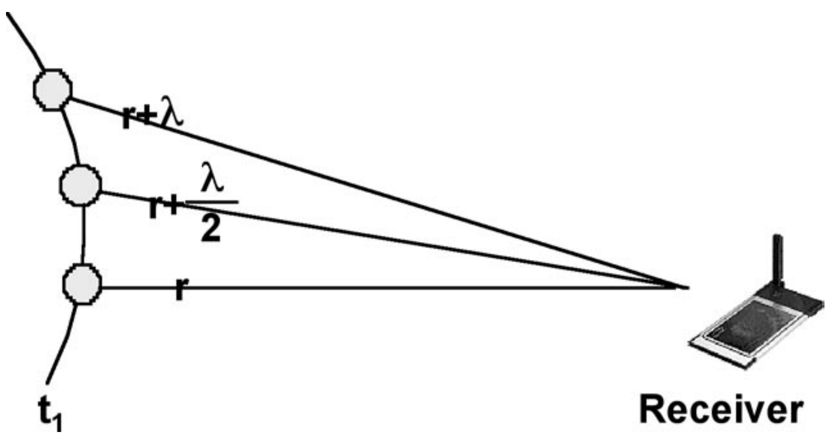


Figure 6.5 The Critical Angle at Which the Wave is 180^{O} Out of Phase

All of the signal coming from the zone of wavelets between the bottom point source and the middle point source are within ½ wavelength of the original. In this zone they are adding together to produce the resultant signal. The entire signal in the zone between the middle point source and the top point source is more than 1800 out of phase with the original and this part of the signal is opposing the first zone. The first zone presents a field to the receiver, the second zone is opposing it. The third zone is, again, in phase with the original and the fourth is, again, opposing. As it turns out, the first zone is the most critical. It's here that the field is within the Rayleigh phase angle. Of course, as has been pointed out, this entire line of reasoning doesn't apply unless there's an obstruction. Without an obstruction the receiver doesn't see any difference in the wavelets, it just sees a spherical wavefront (or the planar presentation of a spherical wavefront). Now let's introduce an obstruction and see what happens. The obstruction is partially blocking the first zone. In this case there are two effects.

First, the wavefront is diffracted downwards, around the edge of the obstruction. Secondly, and more importantly, the opposing phase relationship of the second zone has less "straight-on" signal from the first zone, and the signal power at the receiver is reduced.

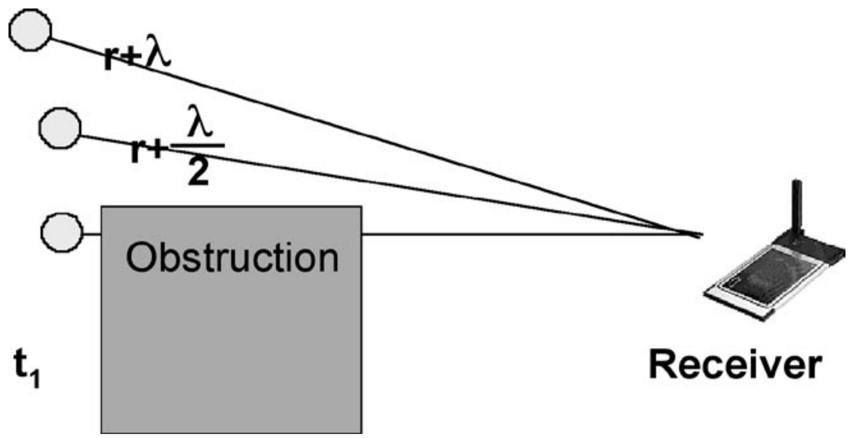


Figure 6.6 The Effect of an Obstruction on the Received Wavelets

With an obstruction present not all of the wavelets from the point sources at t_1 are participating in the construction of the wavefront at the receiver. If we consider a situation where the receiver is moving it can be seen that this effect is directly related to the relationship between the receiver, the obstacle, and the transmitter and not some sort of magical zones in space surrounding the transmitter.

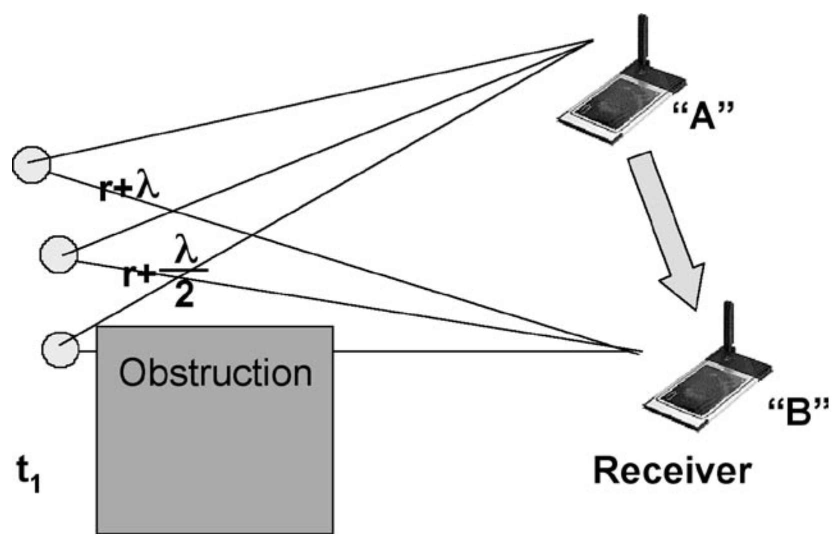


Figure 6.7 The Receiver's Location Determines the Obstructions Affect

When the receiver is at "A" there is no radius of the originally transmitted spherical wavefront that is obstructed. The construction of the top set of three propagation lines in the figure is theoretical, and has no significance to the practical world. As the receiver begins to move downwards in the figure it reaches a point where radial lines from the original transmitting antenna are blocked by the obstruction. At this point the first phase zone begins to be blocked and bad things happen to the signal strength at the receiver. It's very important to realize that these phase zones are not fixed in space relative to the transmitter but, rather, are angular zones related to the direct line-of-sight between transmitter and receiver. As the receiver moves, the line-of-sight line moves as well. It's only when one of the phase zones relative to that line-of-sight is blocked by an obstacle that the phase difference between wavelets impacts a receiver. Remember that it's the first zone that creates the most significant impact on the received signal and degradation due to obstruction in the first zone is exacerbated by the fact that the relative effect of the second zone's opposing influence is magnified when the first zone is blocked.

Fresnel Zones

The zones that have been described are referred to as "Fresnel Zones" and the first Fresnel zone (the inner-most zone) is commonly discussed in 802.11 literature. Unfortunately, there are very few reasonable, non-mathematical descriptions of exactly what a Fresnel zone is, and this leads to confusion. A Fresnel Zone may be thought of as a 3-dimensional oval volume defined along the axes formed by the straight line between a transmitting antenna and a receiving antenna. The figure below shows a single Fresnel Zone volume.

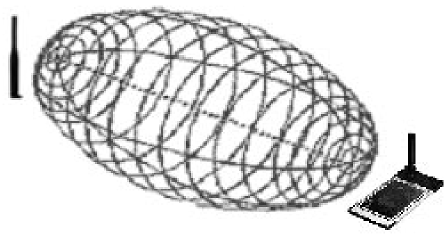


Figure 6.8 The Oval Volume of a Fresnel Zone

Multiple Fresnel Zones are built up around the central line-of-sight axis as shown in the figure below.

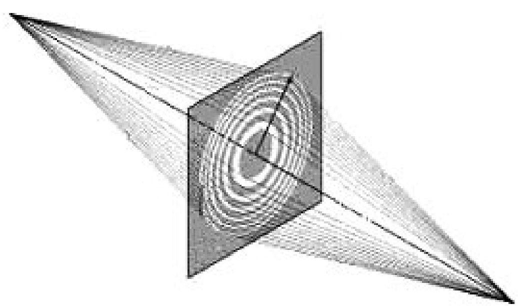


Figure 6.9 Multiple Fresnel Zones Built Up Around the Central Axis

As mentioned, it's only the first Fresnel Zone (the innermost zone) that is of particular significance with regard to signal obstruction in an 802.11 wireless network.

Fresnel Zones are not Related to Antenna Gain or Directivity

The radius of the Fresnel Zone volume at any point along its axis is determined solely by the wavelength of the signal and is completely unrelated to how the energy is dispersed in the field. Therefore, the radius of a Fresnel Zone is unrelated to antenna beamwidth, directivity, or gain.

A directional antenna offers some particular total field energy in some particular direction. That field energy, whatever its value, is either unopposed or opposed by Fresnel Zone interaction when an obstacle gets too close to the straight line-of-sight between the antenna and the receiver. Consider a receiver located to the side of a directional antenna in such a way that it receives 50% of the antenna's output power with an unobstructed line of site. If an obstacle were in the first Fresnel zone surrounding that line of sight then the receiver would get less power. Again, the location of the Fresnel zone is unrelated to the directionality of the antenna. It's only relative to the direct line between transmitter and receiver.

Calculating the Radius of the Fresnel Zones

The radius of the Fresnel zones may be calculated by the following formula.

$$R_n = M\sqrt{\frac{N}{F_{GHz}}\left(\frac{D_1 D_2}{D_1 + D_2}\right)}$$

Where:
R_n is the radius of the nth Fresnel zone
M is a constant of proportionality equal to:
17.3 if R_n is in meters and D_1, D_2 are in kilometers and
72.1 if R_n is in feet and D_1, D_2 are in statute miles
F_{GHz} is the frequency in GHz
N is the Fresnel zone number (Equal to 1 for the 1st Fresnel Zone)
D_1 is the distance from the source to the obstruction
D_2 is the distance from the destination to the obstruction

Obstructions in the First Fresnel Zone

It has been determined that the diffraction characteristics of the first Fresnel zone are the most critical to received signal strength. Calculations have been made concerning the impact of obstructions in the first Fresnel zone and it's been found that the first zone must be at least 60% clear of obstructions. The figure below shows a file cabinet in the Fresnel Zone between an 802.11 access point antenna and a PCMCIA card in a user's notebook computer.

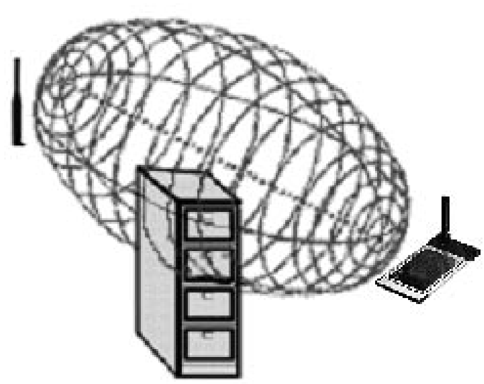


Figure 6.10 Interior Obstructions in the First Fresnel Zone

If this were an outdoor RF implementation and the two antennae were parabolic dishes mounted on buildings across the street from each other, the obstruction could be a tree or another building interposed in the Fresnel Zone.

The first Fresnel Zone must be at least 60% unobstructed. If more than 40% of the zone is blocked it can be assumed that severe degradation of the signal strength will result. This is an important realization. Even though there might be a direct line-of-sight between the wireless adapter in a client computer and the access point on the ceiling, the radius of the first Fresnel zone surrounding the straight line-of-sight path must be taken into consideration. An interior object (like a file cabinet or cubicle wall) may be obstructing the first Fresnel zone and causing unexpected degradation of the signal. When considering outdoor RF implementation this becomes more critical due to the fact that there are fewer opportunities for reflected signals to bounce around in a room and make their way to the receiving antenna. Moreover, outdoor RF typically utilizes directional antennae, further reducing the possibility of multipath reflections.

To assess the clearance in the first Fresnel zone you must first apply the formula to determine the radius of the zone at the distance being assessed. You can consider the maximum radius of the zone that always occurs exactly halfway between the two antennae and use that value as your guideline. From the straight line defined by the direct line-of-sight between the two antennae, remember that the first Fresnel zone must be at least 60% unobstructed. A completely unobstructed zone would, of course, be better!

Practical Examples of the Fresnel Zone Calculation

Let's look at some practical examples using F=2.437 GHz (the center frequency of 802.11b channel 6).

An end-user, sitting in their cubicle, is 50-feet from the access point. What is the radius of the first Fresnel zone at the point where their metal file cabinet is between their notebook computer and the access point? The file cabinet is 4-feet away from the user.

D_1 = 46' from access point
D_2 = 4' from client wireless adapter
First Fresnel Zone = 1.22 feet (1-foot, 2.64-inches)
60% radius = .737 feet (8.44-inches)

For the same end-user, what is the radius of the first Fresnel zone at the point where an upright shelving unit is located halfway between the user and the access point?

D_1 = 25' from access point
D_2 = 25' from client wireless adapter
First Fresnel Zone = 2.26-feet (2-feet, 3.1-inches
60% radius = 1.358-feet (1-foot, 4.3-inches)

What is the maximum radius of the first Fresnel zone when the end-user is 150-feet away from an access point? Remember that the maximum radius occurs halfway between the two antennae.

D_1 = 75' from access point
D_2 = 75' from client wireless adapter
First Fresnel Zone = 3.92-feet (3-feet, 11-inches)
60% radius = 2.35-feet (2-feet, 4.2-inches)

An outdoor 802.11 implementation uses a high-gain antenna to span a 3 statute mile distance between two buildings. What is the maximum radius of the first Fresnel zone (which will be in the middle, between the two antennae)?

D_1 = 1.5 statute miles from antenna #1
D_2 = 1.5 statute miles from antenna #2
First Fresnel Zone = 39.99-feet (39-feet, 11.8-inches)
60% radius = 23.99-feet (23-feet, 11.8-inches)

You can see that the radius required to maintain the 60% unobstructed first Fresnel zone requirement is potentially an issue in an indoor environment but becomes dramatic for outdoor RF. Remember, though, that reflection inside a building will often allow connectivity in spite of a blocked first Fresnel zone. The connectivity may, however, suffer from performance degradation or inconsistency. Remember, too, that the Fresnel zones are not static in space, they are relative to the line-of-sight between the two antennae. Therefore, as a user moves their computer within a building (and the line-of-sight to the access point moves) the first Fresnel zone will move, too. The zone is always relative to the two antennae.

The Fresnel Construction

Consider a transmitter at point P_1 whose spherical field is going to propagate to a receiver at point P_2. An obstruction is in the path at point O. The distance from P_1 to O is d_1 and the distance from O to P_2 is d_2. The path length from A to P_2 is $d_2 + \lambda/2$. Because this path length is ½-wavelength longer than the direct line-of-sight, Huygens wavelets emanating from point A would arrive at P_2 exactly 180^O out of phase with those arriving directly from P_1 (along the d_1 and d_2 line-of-sight). Using this basic diagram as a starting point we're able to derive the radius of the first Fresnel zone from an understanding of the relationships between the sides of the triangles.

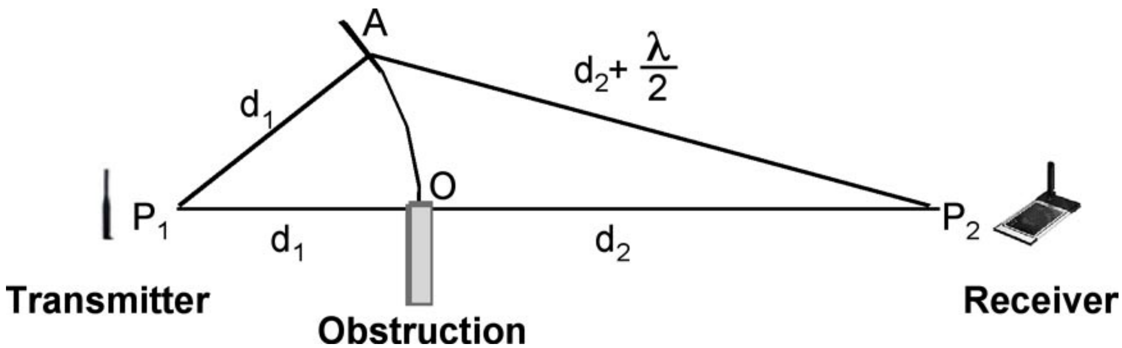


Figure 6.11 The Pythagorean Construction of the First Fresnel Zone

Fresnel used the Pythagorean Theorem to construct an equation for the radius of the first diffraction zone. The Fresnel construction is a beautiful mathematical representation of a physical reality and we'll walk through each step of the derivation. All that's needed to follow the math is some basic algebra and trigonometry. For those readers who have been away from school for too long, you need simply recall two fundamental things from math class:

The Pythagorean Theorem:

$\text{hypotenuse}^2 = \text{base}^2 + \text{height}^2$ for any right triangle on a plane
Squared factors of a polynomial:
$(a+b)^2 = a^2 + 2ab + b^2$ and $(a-b)^2 = a^2 - 2ab + b^2$

We begin by constructing a vertical line downwards from the point (A) where the $d_2+\lambda/2$ line meets the spherical wavefront. This line (AB) has a height h and meets the line-of-sight (P_1P_2) at a distance f away from the point where the spherical wavefront meets the line-of-sight (O). The radius of the first Fresnel zone is, therefore, the distance h from point A to point B.

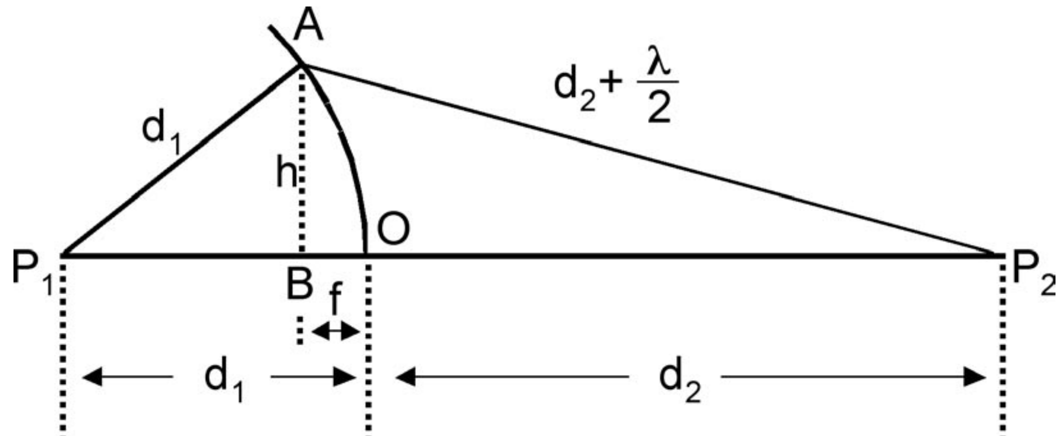


Figure 6.12 Two Triangles Are Constructed Between Transmitter and Receiver

You can see that two triangles are formed in Figure 6.12 (above). On the left is triangle P_1AB. The hypotenuse is d_1 and the height is h. The base of this triangle is the line from P_1 to B which can be seen to have a length of (d_1-f)

On the right is triangle P_2AB. The hypotenuse is $d_2+\lambda/2$. The base of this triangle is the line from P_2 to B which can be seen to have a length of (d_2+f)

The basic Pythagorean relationships for these two triangles are:

$$d_1^2 = h^2 + (d_1 - f)^2 \qquad \left[d_2 + \frac{\lambda}{2}\right]^2 = h^2 + (d_2 + f)^2$$

From these basic relationships we are now able to calculate the radius of the first Fresnel zone by solving the equations for h. When we're done we'll have a usable equation for calculating the first Fresnel zone radius. The number of engineers in the general 802.11 WLAN industry who truly understand the inner-workings of Fresnel's calculations is very small. To bring you into the elite group of those who are "in the know" the derivation is presented in complete detail. Each step of the solution follows.

Step 1: Solve the Pythagorean relationships for the common leg, h:

$$d_1^2 - (d_1 - f)^2 = h^2 \qquad \left[d_2 + \frac{\lambda}{2}\right]^2 - (d_2 + f)^2 = h^2$$

Step 2: Since h is a common leg it follows that the two solutions from Step 1 are equal to each other.

$$d_1^2 - (d_1 - f)^2 = \left[d_2 + \frac{\lambda}{2}\right]^2 - (d_2 + f)^2$$

Step 3: Begin by squaring the three factors:

$$d_1^2 - (d_1^2 - 2d_1 f + f^2) = \left(d_2^2 + \frac{2d_2\lambda}{2} + \frac{\lambda^2}{4}\right) - (d_2^2 + 2d_2 f + f^2)$$

Step 4: The minus signs preceding the quantities in parentheses must be applied to each term within the parentheses:

$$d_1^2 - d_1^2 + 2d_1 f - f^2 = \left(d_2^2 + \frac{2d_2\lambda}{2} + \frac{\lambda^2}{4}\right) - d_2^2 - 2d_2 f - f^2$$

Step 5: Now equal terms can be canceled on the left and right:

$$\cancel{d_1^2} - \cancel{d_1^2} + 2d_1 f - \cancel{f^2} = \left(\cancel{d_2^2} + \frac{\cancel{2}d_2\lambda}{\cancel{2}} + \frac{\lambda^2}{4}\right) - \cancel{d_2^2} - 2d_2 f - \cancel{f^2}$$

Step 6: The resulting equation becomes:

$$2d_1 f = d_2\lambda + \frac{\lambda^2}{4} - 2d_2 f$$

Step 7: By adding $2d_2f$ to both sides of the equation the result becomes:

$$2d_1 f + 2d_2 f = d_2\lambda + \frac{\lambda^2}{4}$$

Step 8: The 2f is now factored out on the left-hand side:

$$2f(d_1 + d_2) = d_2\lambda + \frac{\lambda^2}{4}$$

Step 9: Now both sides of the equation are divided by $2(d_1+d_2)$ which removes the term from the left and solves the equation for f:

$$f = \frac{d_2\lambda + \frac{\lambda^2}{4}}{2\,(d_1 + d_2)}$$

Step 10: Remember that λ is the wavelength, and in the 802.11 bands this value will be less than 1-foot long. For outdoor RF the value of d_1 and d_2 could be many miles. Because λ is so much smaller than d_1 or d_2 a rough (but reasonability accurate) result can be obtained for the equation by simply treating $\lambda^2/4$ as if it were zero. This method is called "taking the first approximation" and is a standard tool in the mathematician's bag of tricks. The first approximation for the equation is:

$$f = \frac{d_2\lambda + \cancelto{0}{\frac{\lambda^2}{4}}}{2\,(d_1 + d_2)} \qquad \frac{\lambda^2}{4} \rightarrow 0$$

The arrow indicates that the term $\lambda^2/4$ has gone to zero in both the equation and in the notation to the right.

Step 11: A solution has now been extracted for the length of f:

$$f = \frac{d_2\lambda}{2\,(d_1 + d_2)}$$

Recall that in Step 1 the height of the right-hand triangle was given by the Pythagorean Theorem as:

$$h^2 = \left(d_2 + \frac{\lambda}{2}\right)^2 - (d_2 + f)^2$$

Step 12: Substitute the extracted equation for f (from Step 11) into the original equation for the right-hand triangle (from Step 1):

$$h^2 = \left(d_2 + \frac{\lambda}{2}\right)^2 - \left(d_2 + \frac{d_2\lambda}{2\,(d_1 + d_2)}\right)^2$$

Step 13: Complete the square of the first term $(d_2+\lambda/2)^2$ in the equation from Step 12:

$$d_2^{\,2} + \frac{2d_2\lambda}{2} + \frac{\lambda^2}{4}$$

Step 14: Cancel the equal factors and take the first approximation to simplify the expression:

$$d_2^{\,2} + \frac{\cancel{2}d_2\lambda}{\cancel{2}} + \cancel{\frac{\lambda^2}{4}} \rightarrow 0 \qquad \frac{\lambda^2}{4} \rightarrow 0$$

Step 15: The simplified first term in the equation from Step 12 is, therefore:

$$d_2^{\,2} + d_2\lambda$$

Step 16: Return to the equation in Step 12 and complete the square in the second term:

$$d_2^{\,2} + \frac{2d_2^{\,2}\lambda}{2\,(d_1 + d_2)} + \frac{d_2^{\,2}\lambda^2}{4\,(d_1 + d_2)^2}$$

Step 17: The 2 factors out on the left and the resulting terms become:

$$d_2^{\,2} + \frac{d_2^{\,2}\lambda}{(d_1 + d_2)} + \frac{d_2^{\,2}\lambda^2}{4\,(d_1 + d_2)^2}$$

Step 18: Notice, on the right, that λ^2 appears in the numerator, and 4 appears in the denominator. As was done before with $\lambda^2/4$, the first approximation goes to zero and, therefore, the entire second term becomes zero and drops off:

$$d_2^2 + \frac{d_2^2\lambda}{(d_1 + d_2)} + \cancel{\frac{d_2^2\lambda^2}{4(d_1 + d_2)^2}} \rightarrow 0 \qquad \frac{\lambda^2}{4} \rightarrow 0$$

Step 19: Having dropped the entire second term in the previous step the entire right-hand term for the equation from Step 12 now becomes:

$$d_2^2 + \frac{d_2^2\lambda}{(d_1 + d_2)}$$

Step 20: The equation from Step 12 may now be rewritten with its terms squared and with the first approximation included. As you work through this step, remember that the second term on the right-hand side of the equation in Step 12 is subtracted from the first term. That's why the sign changes to a minus (as shown now below) between d_2^2 and $d_2^2\lambda/(d_1+d_2)$:

$$h^2 = d_2^2 + d_2\lambda - d_2^2 - \frac{d_2^2\lambda}{(d_1 + d_2)}$$

Step 21: The d_2^2 terms now cancel:

$$h^2 = \cancel{d_2^2} + d_2\lambda - \cancel{d_2^2} - \frac{d_2^2\lambda}{(d_1 + d_2)}$$

Step 22: Continuing, multiplying $d_2\lambda$ by $(d_1+d_2)/(d_1+d_2)$ is simply multiplying it by 1 and doesn't change its value (but it's going to help in the calculation!)

$$h^2 = d_2\lambda \frac{(d_1 + d_2)}{(d_1 + d_2)} - \frac{d_2^2\lambda}{(d_1 + d_2)}$$

Step 23: Perform the multiplication in the first term:

$$h^2 = \frac{d_1 d_2 \lambda + d_2^2 \lambda}{(d_1 + d_2)} - \frac{d_2^2 \lambda}{(d_1 + d_2)}$$

Step 24: Now the denominators of the two terms are the same and the numerators may be combined:

$$h^2 = \frac{d_1 d_2 \lambda + d_2^2 \lambda - d_2^2 \lambda}{(d_1 + d_2)}$$

Step 25: Cancel the terms that come out due to subtraction in the numerator:

$$h^2 = \frac{d_1 d_2 \lambda + \cancel{d_2^2 \lambda} - \cancel{d_2^2 \lambda}}{(d_1 + d_2)}$$

Step 26: The resulting equation now becomes:

$$h^2 = \frac{d_1 d_2 \lambda}{(d_1 + d_2)}$$

Step 27: Solve this equation for h to yield the first final form of the equation:

$$h = \sqrt{\lambda \frac{d_1 d_2}{(d_1 + d_2)}}$$

This is the general equation for the radius of the first Fresnel zone, derived through the Pythagorean Theorem from the construction of the triangles formed by representing the radius of the spherical wavefront and the $d_2+\lambda/2$ length to the receiver.

Dealing with an Unfriendly Equation

The general equation for the radius of the first Fresnel zone is, unfortunately, not particularly friendly in practical terms. The first unfriendly aspect is that all of the variables (h, d_1, d_2, and λ) must use the same unit of distance measurement. In practice the radius of the first Fresnel zone is most critical in outdoor RF where it would be most convenient to use feet to measure the radius (h) and miles to measure the distance from the transmitter to the obstruction (d_1) and from the obstruction to the receiver (d_2). The second unfriendly aspect is that most specifications for 802.11 speak in terms of frequency (2.4 GHz, 5.8 GHz) and not in terms of wavelength (λ).

To make the Fresnel zone equation more usable in the field it is necessary to apply some conversion factors for miles and frequency. The introduction of the conversion factors results in a value, called the *constant of proportionality*, which is used to calculate the zone radius with mixed units and with frequency instead of wavelength. The derivation of this constant of proportionality follows.

Step 1: We'll first convert wavelength to frequency by remembering that they are related by the speed of light so that $\lambda = c/F$. Substituting c/F for λ results in the following:

$$h = \sqrt{\frac{c}{F} \frac{d_1 d_2}{(d_1 + d_2)}}$$

Step 2: We want the radius to be measured in feet and the distances from transmitter and receiver to the obstacle to be measured in miles. Every time there's a variable that's going to be given in miles, we must convert it to feet by multiplying by 5280 (feet in one statute mile), as follows:

$$h = \sqrt{\frac{c}{F} \frac{(d_1 * 5280)(d_2 * 5280)}{(d_1 * 5280) + (d_2 * 5280)}}$$

Step 3: The denominator under the root can now be factored as follows:

$$h = \sqrt{\frac{c}{F} \frac{(d_1 * 5280)(d_2 * 5280)}{(d_1 + d_2) * 5280}}$$

Step 4: Like terms in the numerator and denominator now cancel:

$$h = \sqrt{\frac{c}{F}\,\frac{(d_1 * 5280)(d_2 * \cancel{5280})}{(d_1 + d_2) * \cancel{5280}}}$$

Step 5: The resulting equation at this step now becomes:

$$h = \sqrt{\frac{c * 5280 * d_1 d_2}{F * (d_1 + d_2)}}$$

Step 6: There's another unfriendly aspect to consider now. The basic relationship between wavelength and frequency ($F=c/\lambda$) uses cycles/second to measure F. One cycle/second is also called one Hertz. Wireless networks (and many other environments) typically measure frequency in Gigahertz (GHz) where 1 GHz = 1,000,000,000 Hertz. One Hertz is, therefore, 1/1,000,000,000 GHz ($1/10^{-9}$ GHz) and this conversion factor must be included in the equation as well:

$$h = \sqrt{\frac{c * 10^{-9} * 5280 * d_1 d_2}{F * (d_1 + d_2)}}$$

Step 7: The constant terms (c, 10^{-9}, and 5280) can be grouped and separated from the rest of the equation as follows:

$$h = \sqrt{c * 10^{-9} * 5280} * \sqrt{\frac{d_1 d_2}{F * (d_1 + d_2)}}$$

Step 8: The speed of light (represented by the lower-case letter "c") is typically stated as being 300,000,000 meters/second in a vacuum. The most contemporary measurements for the speed of light put it slightly less than this value, with 299,792,458 meters/second being the value given in scientific literature. For the purposes of the calculations that follow, however, the typical value will be used. As it turns out, the most widely published equations for dealing with the Fresnel zones use 300,000,000 meters/second in their inner workings and so, to obtain results that are consistent with the ones you'll most likely encounter in the 802.11 wireless networking realm, this value will be given to c. There are 1609 meters in one mile and, therefore, by dividing, it can be found that the speed of light may be represented as 186,404.8714 miles/second. Moreover, there are 5280 feet in a statute mile and so, by multiplying, the speed of light could also be given as 984,217,720.8898 feet/second. These values will be important because c is included in the constant of proportionality while calculating the radius of the first Fresnel zone. If the radius is going to be measured in feet then the speed of light will have to be represented in feet/second. If the radius is going to be measured in meters then the speed of light is going to have to be represented in meters/second. In every case the units must all match, feet-for-feet, miles-for-miles, meters-for-meters, and so forth. All distances must be represented using the same length units.

Remember that the units under the left-hand root are now feet, and the units under the right-hand square root are statute miles. To calculate the left-hand square root it will be necessary to represent c in feet/second. Using the value 984,217,720.8898 feet/second for c, the square root of the product $c*10^{-9}*5280$ = 72.08792941. This value for the constant of proportionality is commonly rounded up to 72.1 and the equation for the radius of the first Fresnel zone becomes:

$$h = 72.1\sqrt{\frac{d_1 d_2}{F * (d_1 + d_2)}}$$

One More Equation

In literature you will encounter the equation for the maximum radius of the first Fresnel zone when the length of the path between transmitter and receiver is known. For your edification, the derivation is shown here.

Step 1: Start with the general equation for the radius of the first Fresnel zone and assume that the obstacle is in the exact center of the transmission path. At this point the radius of the zone is at its maximum. At this point in the path $d_1 = d_2$ and the length of the path (L) is equal to d_1+d_2. Consequently:

$$d_1 = d_2 = 1/2\ L$$

Step 2: Substitute these values into the general equation:

$$h = 72.1\sqrt{\frac{1/2\ L * \ 1/2}{F * L}}$$

Step 3: Multiply the fractions in the numerator:

$$h = 72.1\sqrt{\frac{1/4 * L * L}{F * L}}$$

Step 4: Cancel the matching terms in the numerator and denominator and multiply by 4/4 to get the fraction out of the numerator:

$$h = 72.1\sqrt{\frac{1/4 * L * \cancel{L}}{F * \cancel{L}} * \frac{4}{4}}$$

Step 5: The result is the equation for the maximum first Fresnel zone radius (in feet) when the distance between transmitter and receiver (L) is given in statute miles and the frequency (F) is given in GHz.

$$h = 72.1\sqrt{\frac{L}{4F}}$$

The Erroneous Constant of Proportionality

It's been shown, in excruciating detail, how the constant of proportionality for calculating the radius of the first Fresnel Zone is equal to 72.1. We've also discussed that the first Fresnel zone must be 60% unobstructed for proper signal reception. Sometimes you'll see the Fresnel Zone formula presented with the constant of proportionality already adjusted to 60% of its proper value. This is a simple modification to the basic equation since 72.1 * .60 = 43.26, which is rounded up to 43.3, as shown in Figure 6.13, below.

$$h_{60\%} = 43.3\sqrt{\frac{d_1 d_2}{F * (d_1 + d_2)}} \qquad h_{60\%} = 43.3\sqrt{\frac{L}{4F}}$$

Figure 6.13 The Typical Presentations of the Fresnel Zone Equations

Occasionally you will encounter an erroneous use of the constant of proportionality. A discussion of the radius of the first Fresnel zone will present the equations (using 43.3) for calculating the full size of the zone, and the accompanying literature will explain that this radius must be 60% unobstructed! Using the erroneous explanation results in calculating an unobstructed zone requirement that is 60% of 60% of the full first Fresnel zone radius, and not simply 60%. It's hoped that the derivation of the Fresnel zone equations provided here will enlighten those who have erred in their understanding of the math. It's probably the case that someone's misunderstanding was taught to someone else who believed it to be factual. Then the misinformation spread as one writer, speaker, or technical instructor trusted that the equations they had learned were accurate. You now know the truth!

Concluding Thoughts

The mysterious nature of electromagnetic wave propagation and the practical application of 802.11 wireless networking principles come together every day for the WLAN designer, support engineer, or network manager. Hopefully this somewhat lengthy discussion has shed some light on how it all works. Hopefully, too, you're motivated and challenged to do some self-study, web research, or even sign up for a class at a local college to explore RF engineering in greater detail.

Albert Einstein was asked to explain how the wireless telegraph worked. He replied "The wireless telegraph is not difficult to understand. The ordinary telegraph is like a very long cat - you pull the tail in New York, and it meows in Los Angeles. The wireless telegraph is the same, only without the cat."

Appendix A

The Solution To Zeno's and Bardwell's Paradoxes

The flaw in both Zeno's paradoxes and Bardwell's ERP Paradox relates to the fact that the mathematical solution proposed (an iterative series of calculations) requires that an infinite series of numbers be considered. If that were true then it would take Achilles an infinite amount of time to reach the turtle and power at an antenna would be infinite. The rules of calculus tell us, however, that the problem is not one of an infinite series but, rather, of a value that approaches a limit. The false argument proposes that there are an infinite series of time-points between two events. Then, to create the paradox, the stipulation is that an infinite series of points can't be considered in a finite amount of time. This is an invalid approach since the supposedly infinite series of calculations takes place on a bounded range (access point antenna to measurement point) that is finite in length. The question posed by the paradox is constructed within a finite set of boundaries. In the case of an 802.11 transmission, the field producing the Effective Radiated Power begins at the antenna (for the near field) or at the fuzzy boundary between the reactive near field and the far field (for far field measurements). At the starting point for ERP the power in the field is determined by the input power presented across the impedance (resistance) of the antenna element, or elements. Since (simplistically stated), power equals voltage times amperage, and amperage equals voltage divided by resistance (Ohm's law) the input power to the antenna's radiating element can be calculated. After that, the near field power decreases proportionally to the cube of the distance and the far field power decreases proportionally to the square of the distance. The answer to Bardwell's ERP Paradox is, succinctly stated "The initial power output of an antenna is calculated based on the input power and the distance from the antenna to any measurement point is finite. The fact that this finite distance contains an infinite number of mathematical 'halves' is of no consequence to Maxwell's equations or to the basic equations for field strength based on point charges."

The infinite series that underlies both Zeno's and Bardwell's paradoxes are applicable only in the realm of classical physics which, of course, is where the 802.11 WLAN engineer will spend their time. Interestingly, there are actually aspects of Bardwell's Paradox that do remain unsolved even by contemporary Quantum Electrodynamics (QED). The quantum physicist would be able to explain that there are issues relating to the field strength diverging that remain a mystery even today!

Appendix B

Trigonometric Relationships: Tangent, Sine, and Cosine

Consider a triangle ABC such that AB forms the hypotenuse, AC lies along the x-axis, and CB represents the y-axis height, as shown in Figure B.1, Diagram 1 on the left, below.

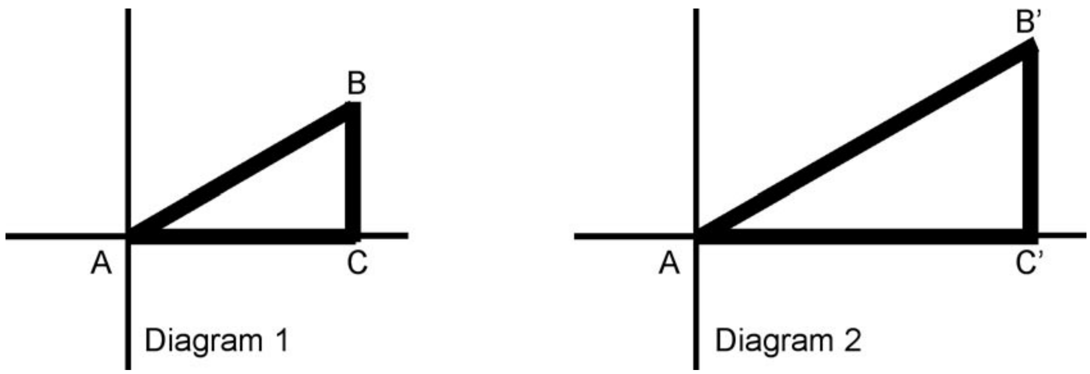


Figure B.1: Trigonometric Relationships In Right Triangles

In Figure B.1 Diagram 2 (above) the BC line has moved to the right (represented now as B’ and C’). There is an immutable ratio between the length of AC and the length of BC. As AC gets longer, BC gets taller, but always in a fixed ratio; a ratio based on the size of the angle at A. The ratio between the x-axis value and the y-axis value (AC and BC) is called the *tangent* of angle A. The basic trigonometric relationship is that the tangent value is equal to the ratio of the length of the side opposite the angle (BC) divided by the length of the side adjacent to the angle (AC). This gives rise to two additional relationships that are the sine and cosine ratios. The sine is the ratio of the length of the side opposite the angle (BC) divided by the length of the hypotenuse (AB). Cosine is the ratio of the length of the side adjacent the angle (AC) divided by the length of the hypotenuse (AB). These relationships are summarized in Figure B.2 below.

$$\tan = \frac{\text{Opposite}}{\text{Adjacent}} \qquad \sin = \frac{\text{Opposite}}{\text{Hypotenuse}} \qquad \cos = \frac{\text{Adjacent}}{\text{Hypotenuse}}$$

Figure B.2: The Basic Trigonometric Relationships in a Right Triangle

Appendix C

Representational Systems for Vector Description

Mathematicians have developed more than a dozen ways of representing the special relationships between physical things. Three of these that are often encountered are through the use of coordinate axes, cylindrical coordinates, and spherical coordinates.

A street map is a two dimensional representation of a town and, like the simple coordinate plane (with an X and a Y axis) that you may have used in school, any location can be represented with a horizontal (X) and vertical (Y) value. A Z-axis can be added to the picture so that any point in space can be identified as P(x,y,z), where the three coordinates are plotted on three axes of the coordinate plane. When you see strange formulae in some antenna book or in some web article discussing electromagnetism (or anything else, for that matter) and you see that 3-dimensional coordinates are presented as an x, y, and z value, you know you're using the classic coordinate axes method of representation.

Electronics (and many other disciplines) often involves evaluation of a plane that's wrapped into a circle (a cylinder), like a piece of wire. A second representational system has been developed for solving equations involving a cylinder. There's a number and an angle that represents the distance up or down the cylinder at which the point being referenced is located and then the angle based on some 0^O on the circle that forms the end of the cylinder. This is called a cylindrical representation (…no surprise there!). When you see an equation where position is represented by two numbers and an angle you're working with a cylindrical format, as shown below.

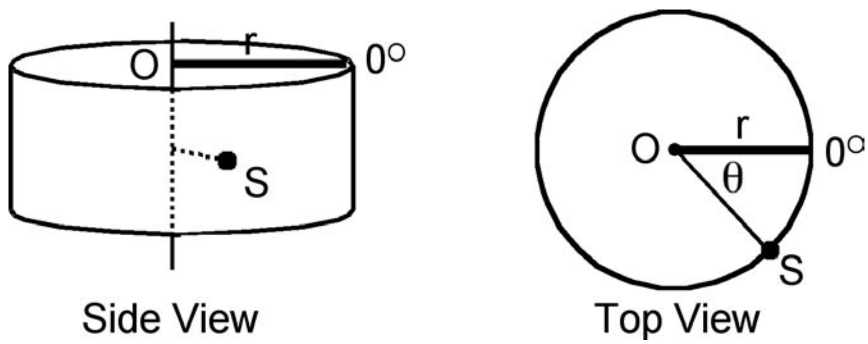


Figure C.1 Vectors Represented Using Cylindrical Coordinates

On the left (Figure C.1 above) is the side view of the representation. A line with length r extends from the vertical line in the middle of the cylinder (at point O) to the circumference of the cylinder. This line has been defined as the 0^O line. Below the radial line at O is another line, meeting the circumference of the cylinder at point S. The angle (clearly seen in the top view) between the two lines is shown as θ. A vector represented in cylindrical form would be written $S(y,r,\theta)$, giving the point on the vertical y-axis for the origin of the vector, the radius of the cylinder, and the angle formed relative to 0^O.

Ultimately, equations often involve points that are naturally in spherical relationships to one-another (like the energy points in a spherically expanding wavefront). The spherical coordinate system (widely used in dealing with electromagnetic radiation equations) is represented below in Figure C.2.

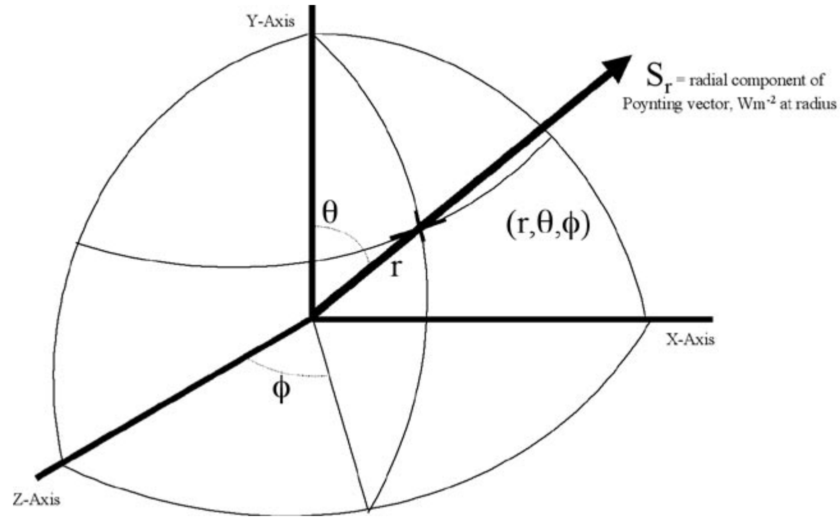


Figure C.2 The Spherical Coordinate System

As you examine the spherical coordinate system figure (above) you see that a particular direction (S_r, the radial component of a Poynting vector in units of Watts per square meter) is shown emanating from the point where the X-, Y-, and Z-axis meet in the center. Instead of representing a point on S_r with an x, y, and z coordinate the angle from the x-axis towards the y-axis is given as θ (theta) and the angle from the x-axis towards the z-axis is given as ϕ (rho). A point at distance r from the center is represented as (r, θ,ϕ).

Appendix D

Electromagnetic Forces at the Quantum Level

In 1900, Max Planck presented a theory on the quantization of energy levels of thermal radiation and this is considered the birth of quantum physics. In 1905, Albert Einstein postulated that all energy is quantized, including light and electromagnetism. The photon, the quanta of electromagnetic energy, can be considered to have characteristics that are like a localized particle, and characteristics that are like a wave. Quantum physics stipulates that any energy must consist of individual packets, or quanta, and that stipulation can be described quite clearly when the field is in motion and energy is changing. The statement implies, however, that even static (unchanging) electromagnetic fields must consist of particles. These particles are called virtual photons.

The core concept behind energy quanta is the idea that all electromagnetic energy is transferred in integer quantities of a fundamental unit. This unit is called Planck's constant and it's represented by the letter h. Another principle of quantum physics is the Heisenberg uncertainty principle. It states that you can never know both the exact position and momentum of any particle and bounds the errors in determining both in the expression $\Delta p \Delta x \geq h/4\pi$ where Δp is the momentum and Δx is the position. The uncertainty principle is not a limit set by the accuracy of measuring equipment. It is a fundamental property of nature.

The uncertainty principle can also be used to conclude that particles of small enough energy and short enough life spans can exist, but you can never measure them. This idea can be stated by the expression $\Delta E \Delta t \leq h/4\pi$ where ΔE is the energy and Δt is the life time. This expression allows for *virtual photons* which can spontaneously appear and disappear as long as they obey both the uncertainty principle and the law of conservation of energy.

The stored energy in the near field of an electromagnetic radiator allows the creation of virtual photons. They carry the electromagnetic force in the non-radiating portion of the field. The virtual photons have all the properties of real photons that make up the radiating fields except that they exist for a very short time and they cease to exist if their source is no longer present.

Appendix E

Enhanced Bibliography

This bibliography is "enhanced" by including not only references to other texts but a brief description of the significance of some of these references. All of these are in the author's personal library and were significant in the preparation of "I'm Going to Let My Chauffeur Answer That".

There are two significant books that are recommended for the library of every networking professional:

Troubleshooting Campus Networks
Priscilla Oppenheimer and Joseph Bardwell
Wiley Publishing 2002 ISBN: 0-471-21013-7

This book (co-authored by Joseph Bardwell, the author of this paper) discusses network protocol behavior and details the interactions between Ethernet, 802.11 wireless, TCP/IP, and the Windows, AppleTalk, Novell, and other upper-layer protocols. If you use a network protocol analyzer this book is an excellent source of information about how to interpret the analyzer's output. The chapter on 802.11 wireless provides a solid understanding of core wireless networking technology.

Deploying License-Free Wireless Wide-Area Networks
Jack Unger
Cisco Press 2003 ISBN 1-58705-069-2

The title is, perhaps, more restrictive than the content reveals. The text compares and contrasts wired and wireless technology and includes excellent explanations of general RF engineering. A comprehensive discussion of antenna types, signal corruption, and network design is applicable to indoor, private 802.11 as well as outdoor RF. Each of the wireless standards (including 802.11) are explained and the switching and routing that must be designed to support wireless networking is also covered. There is a chapter that explains, in great detail, the things that must be considered when conducting an RF site survey.

Additional references that will be useful in your study of RF engineering include the following:

Practical Antenna Handbook
Joseph J. Carr
McGraw-Hill Companies 2001 ISBN: 0-07-137435-3

The Practical Antenna Handbook is considered by many to be the single most important reference guide for antenna theory and design. The opening chapters present an excellent, in-depth discussion of electromagnetic theory relative to the radiating aspects of an antenna element. Only a small portion of the text is directly applicable to the 802.11 environment since the target audience was originally RF engineers working with broadcast radio. There's a good deal of math to allow an engineer to properly design and specify antenna systems.

Antennas for all Applications
John D. Kraus and Ronald J. Marhefka
McGraw-Hill Companies 2002 ISBN: 0-07-232103-2

Originally printed in 1950 and now in it's third edition this book discusses the history of antenna design and provides a solid understanding of wave propagation and related electromagnetic theory. The text is heavily centered on the formulae involved in expressing the antenna theory relevant to each type of antenna so this book is not for those who are intimidated by cosines, square roots, and logarithms. On the other hand, the authors generally stay away from the calculus that lurks just behind the scenes in the discussions.

Wireless Network Performance Handbook
Clint Smith and Curt Gervelis
McGraw Hill 2003 ISBN: 0-07-140655-7

Because this book focuses primarily on wireless networks from the perspective of the cellular telephone provider it has limited value if you're only interested in 802.11 data networking. On the other hand, the discussion of network performance can be extrapolated to the 802.11 data realm and the text includes a perspective on not only 802.11 but competing standards like Bluetooth in the voice and data environment. The protocols used in cell phone communication are explained and the RF site survey and performance assessment issues, including statistical analysis of link quality and capacity, are detailed.

The Feynman Lectures on Physics
Addison-Wesley Publishing 1989 (by California Institute of Technology)
ISBN: 0-201-51003 (v.1)

Dr. Feynman's lectures have been transcribed with notes and drawings and are, without a doubt, the most widely recognized source of physics insight available. The complete 3-volume series is well worth the price and the discussions of electromagnetic theory are a prerequisite to really understanding anything else about RF, 802.11, antennas, or physics in general. If you aspire to a true, in-depth understanding of the physics that underlie 802.11, these volumes are the place to begin.

Electromagnetics Explained
Ron Schmitt
Elsevier Science Publishing 2002 ISBN: 0-7506-7403-2

Here you'll find very concise explanations of the electromagnetic field's characteristics with the accompanying math, but not to the exclusion of a solid written explanation. The text explains the magnetic vector potential, electrical and magnetic strength in antenna propagation patterns, diffraction, and signal corruption. Many of the topics are tied back to their relationship to quantum mechanics. This book is a good all-around read, nicely general in most cases but specific and detailed when the need arises.

Electromagnetic Fields and Interactions
Richard Becker
Blaisdell Publishing Company 1964 ISBN: 0-486-64290-9

Dover Publications (New York) has continued to republish this classic work that describes radiation fields and energy relationships using the equations of electrodynamics. The author provides an excellent introduction to vectors, vector fields, and tensors as applied to the study of electromagnetic fields. To fully appreciate this text the reader must have a solid foundation in advanced calculus since integrals, surface integrals, and partial differentiation appear constantly throughout the book.

Maxwell on the Electromagnetic Field – A Guided Study
Thomas K. Simpson
Rutgers University Press 2001 ISBN: 0-8135-2362 and 0-8135-2363

In this book, Simpson has reproduced James Maxwell's original papers on electromagnetic fields. He then explains and annotates the original text in contemporary terms. The book is very readable; explaining concepts with only the minimum level of mathematics and calculus. It is, however, completely theoretical and serves as a perspective for further study, as opposed to a practical dissertation.

Fresnel Zones in Wireless Links, Zone Plate Lenses, and Antennas
Hristo D. Hristov
Artech House 2000 ISBN: 0-89006-849-6

Advanced calculus is used throughout this book however the mathematically shy reader will be able to gain some very insightful perspective on wave behavior in the first two chapters. An in-depth explanation of Fresnel Zone theory is presented in mathematical terms. This book is definitely worth paging through if

you discover it at your local bookstore but, unless you are a serious RF engineer, the content will be beyond your needs.

Math Refresher for Scientists and Engineers
John R. Fanchi
John Wiley & Sons 2000 ISBN: 0-471-38457-1

If you studied trig and calculus in school but you haven't used it in quite some time, this book will serve as an excellent way to get back up to speed so you'll be able to tackle the RF equations in other texts. Of course, the assumption is that this is a "refresher" and not your first introduction to the topics!

What is quantum mechanics? A physics adventure
Language research foundation 1997 ISBN: 0-9643504-1-6

If quantum mechanics is intriguing then this is probably the best technically-oriented introduction that you'll find. The style of the book is unique and the complex topics are introduced in a way that every reader will gain appreciation for the topic. You won't need a lot of math, although the math that is involved is explained and you won't need to be a physics major either. You're sure to come away from this book with a surprising understanding of quantum mechanics!

QED – The Strange Theory of Light and Matter
Richard P. Feynman
Princeton University Press 1988 ISBN: 0-691-02417-0

Dr. Feynman, in his typical style, mixes insight with science as he explains Quantum Electrodynamics (QED). In this book the reader will be presented with a discussion of photons as both particles and waves and how electrons interact. This book does NOT require any advanced math and is exceptionally readable.

The Picture Book of Quantum Mechanics
Siegmund Brandt
Wiley 2001 ISBN: 0-387-95141-5

The author has provided 3-dimensional graphic representations of quantum probability fields and explains, with appropriate mathematical references, the concepts of probability amplitude. This text provides an in-depth explanation of the "wave packet" and, as such, lays the groundwork for understanding the quantum nature of electromagnetic wave propagation. There's a lot of calculus between the covers but the graphic representations of quantum fields are (as they say about pictures) "worth a thousand words".