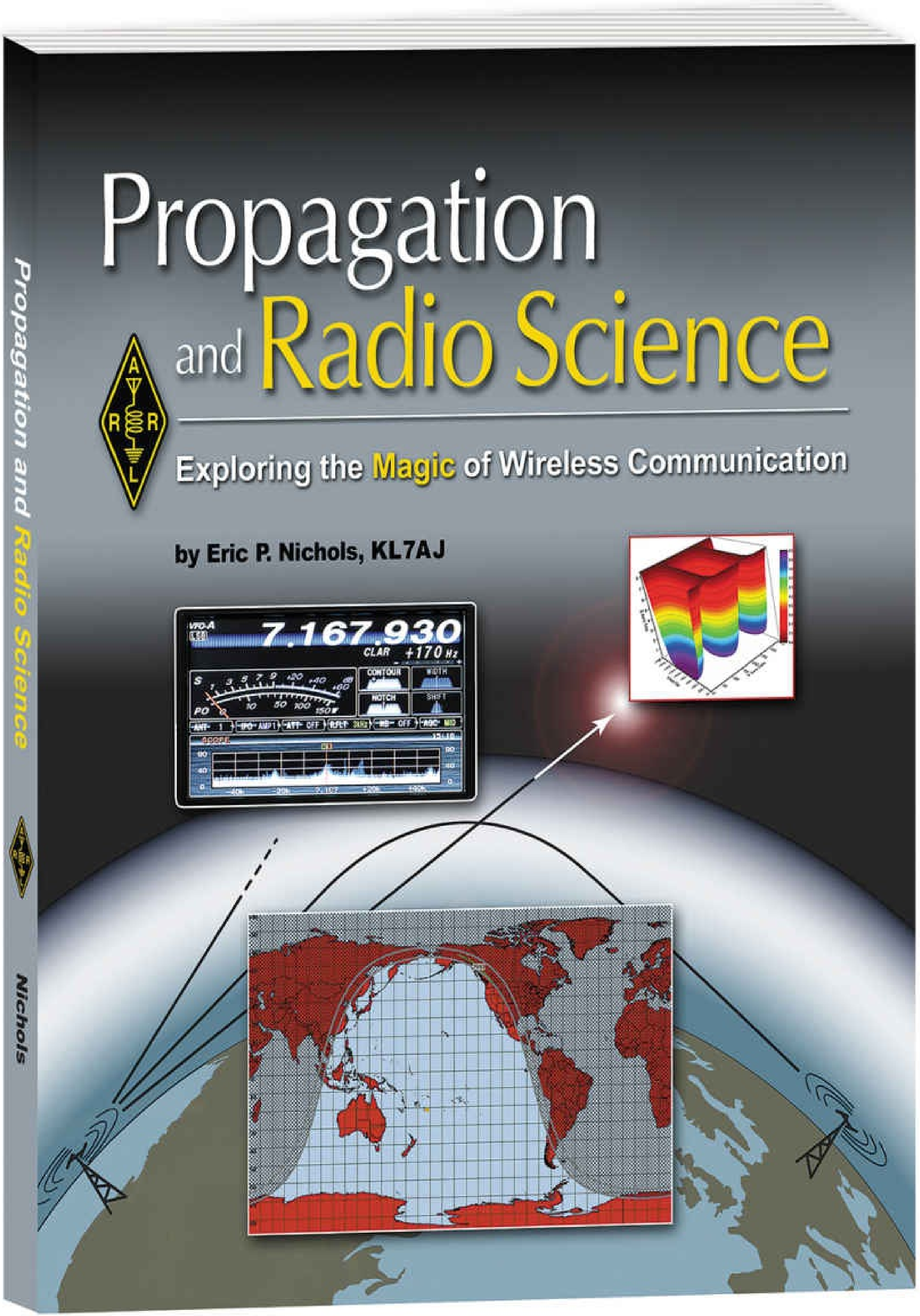
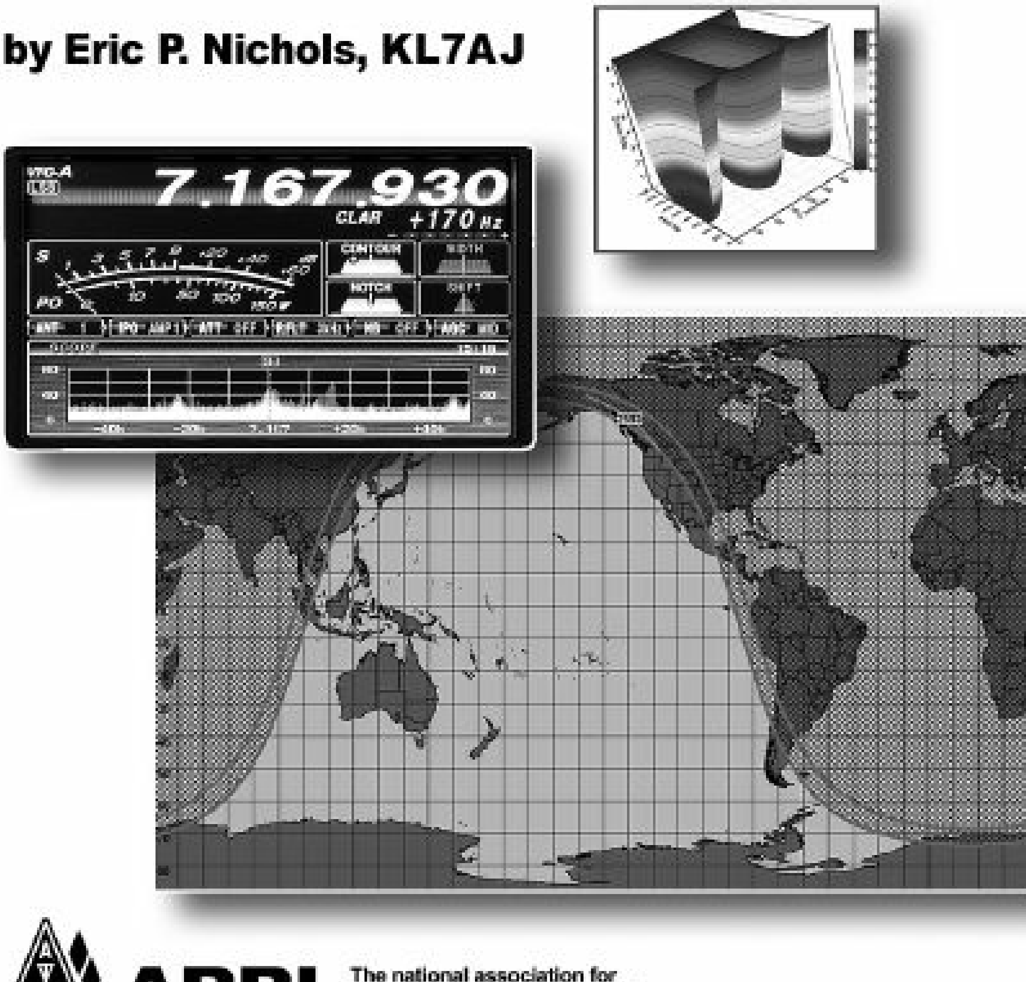

Propagation
and Radio Science
Exploring the Magic of Wireless Communication
by Eric P. Nichols, KL7AJ
7.167.930
CLAR +170 Hz
Propagation and Radio Science
Nichols

Propagation and Radio Science

Exploring the Magic of Wireless Communication

by Eric P. Nichols, KL7AJ



ARRL The national association for AMATEUR RADIO®

Production Staff: Michelle Bloom, WB1ENT, Production Supervisor, Layout
Jodi Morin, KA1JPA, Assistant Production Supervisor, Layout
Sue Fagan, KB1OKW, Graphic Design Supervisor, Cover Design
David Pingree, N1NAS, Senior Technical Illustrator

Printed in the USA

ISBN: 978-1-62595-027-7

First Edition

We strive to produce books without errors. Sometimes mistakes do occur, however. When we become aware of problems in our books (other than obvious typographical errors), we post corrections on the ARRL website. If you think you have found an error, please check **www.arrl.org/notes** for corrections. If you don't find a correction there, please let us know by sending e-mail to **pubsfdbk@arrl.org**.

eBooks created by www.ebookconversion.com

Contents

Foreword

One of the most fascinating and rewarding activities in Amateur Radio is studying the often baffling and sometimes frustrating phenomenon of radio propagation. Actually, radio propagation is not just one phenomenon, but rather *many* phenomena, all of which add up to an intriguing journey of radio waves from transmitter to receiver through the voids of space.

Unlike sound, radio waves travel without the need for a *medium*, a truly amazing property that has intrigued physicists and philosophers for over a century. Understanding how radio waves interact with their environment is an important factor in achieving effective Amateur Radio communications. Amateur frequencies span the entire electromagnetic spectrum, at least in chunks, and each portion of the radio spectrum has unique advantages and potential pitfalls.

A dedicated and skillful radio amateur may take great pains to assure that the station's transmitter, receiver, and antennas are functioning at their optimum potential, yet once a radio wave leaves the antenna, it is at the mercy of the ether.

Because we have no control over the ether, our only option is to understand what it is doing and accommodate its often fickle behavior. This can be a ham's greatest joy or a source of great frustration. And yet, as unpredictable as propagation can be at times, there is a great deal that we do know. And that is what this book is about.

In *Propagation and Radio Science*, author Eric Nichols, KL7AJ, puts radio propagation front and center. In his lively and engaging writing style, Eric explains many of the phenomena we observe on the amateur bands and encourages us to learn more and experiment using readily available tools. We hope you will make use of the wealth of information found here.

David Sumner, K1ZZ
Chief Executive Officer
Newington, Connecticut
March 2015

Preface

The Invisible Journey

For about a century, radio amateurs have been launching radio signals into the ether, using equipment ranging from the primitive to the elegant. Technology and technique has improved enormously over the years, greatly increasing the reliability and effectiveness of this thing we call radio. However, the one thing radio amateurs have no control over is *propagation*, the unseen path between transmitter and receiver that every radio signal must take, whether across the street or to the Moon and back. There are countless varieties of ways a radio signal can get from Point A to Point B, a path that can be fascinating, frustrating, mystifying, and mesmerizing.

There's a lot we know about radio propagation, but there's a lot more we *don't* know. It is a field still wide open to exploration and experimentation. And for radio amateurs, the ether is still *absolutely free*.

Many excellent Amateur Radio books include valuable sections on radio propagation, but the subject is so vast and wide that a chapter or two simply cannot do justice to this most fascinating aspect of Amateur Radio. It is appropriate to have a book dedicated to *nothing but* radio propagation. *Propagation and Radio Science* is just that book; in this comprehensive tome we explore radio propagation from dc to daylight, from the trivial to the truly bizarre. We trust you will find it a worthy companion to the venerable *ARRL Handbook* and *ARRL Antenna Book*.

In *Propagation and Radio Science*, the most daunting mathematics is "safely" tucked away, while the free-flowing narrative presents the story of radio propagation in simple, intuitive, and graphic terms. This method will allow the reader to understand this somewhat complex topic on different levels, according to his or her experience and needs, either practical application, or a complete theoretical exploration. In the final chapters, we will delve into the still-unknown nooks and crannies of radio propagation and suggest areas of exploration for the curious radio amateur.

Special thanks to Carl Leutzelschwab, K9LA, for reviewing this work.

Eric P. Nichols, KL7AJ

North Pole Alaska, March 2015

A Wide World for the Technician Class Amateur

There's a common misconception that Technician class Amateur Radio operators don't have high frequency (HF) privileges, and must wait until they upgrade to General or Amateur Extra status to take advantage of worldwide shortwave communications. This is not true. As a Technician licensee, you have "inherited" all the HF privileges of the former *Novice class* license, which affords privileges on 80, 40, 15, and 10 meters. This is in addition to the more commonly advertised VHF and UHF privileges you already know about.

Back in the day (before the codeless Technician boom in the early 1990s), just about every new radio amateur started out with a Novice license. When the Novice license was created in 1951, Novices had some HF CW privileges, along with CW and voice on 2 meters. The license was good for two years and not renewable. In later years, the 2 meter privileges went away and some 10 meter voice and data privileges were added and the Novice license was renewable with the same terms as the other licenses.

While restricted to CW operation (and later a small segment on 10 meter RTTY, data and SSB) and relatively low power, most "old timers" worked the world on the Novice bands. By the time they upgraded to one of the more advanced licenses, most hams had a good feel for HF propagation. In fact, long before obtaining their Novice licenses, most hams had some experience tuning around on the shortwave bands with a general coverage receiver (SWLing) and had at least an idea of what to expect when it came to long distance radio behavior.

This is no longer the case. A lot of recent Amateur Radio training material geared toward the new Technician license mentions HF as a mere footnote, if at all, while concentrating on local VHF and UHF operation. Most new hams have little or no idea what HF propagation is all about; a fairly large proportion of them have never operated a shortwave radio, even briefly. This is a shame, because much of the exciting stuff happening in Amateur Radio is on the HF bands.

We'll take some considerable time here to talk about what you can do on HF with "only" a Technician license, so you have no excuse not to explore the core of Amateur Radio.

The Code is the Mode

Except for 10 meters, where a portion of the band is allocated to voice (phone) privileges, your HF Technician privileges allow for CW (Morse code) only. While there is no prohibition against using a computer to generate and receive CW, there is no reason not to learn Morse code (more officially known as the International Radiotelegraph Code). Ironically enough, now that Morse code is not required for any class of license, there is actually *more* high quality CW on the air than ever before. This is because a lot of hams learn CW because they *want* to, not because they have to!

Besides being a great deal of fun, CW can punch through poor radio conditions quite well, and it uses *very* little bandwidth. In other words, you can fit a lot of CW stations within a limited frequency range.

A Good Sampling

While the frequency ranges of each HF Technician band are rather small, the four HF bands are nicely distributed throughout the entire HF Amateur Radio frequency range, giving you a wonderful sampling of each type of HF radio propagation. And there *are* vast differences in propagation characteristics between each Technician HF band. We'll go into these different bands in detail so you can take full advantage of your HF privileges, whether you decide to upgrade or remain a Tech indefinitely.

80 Meters: Your Friendly Local Ham Band

The 80 meter band (3.525 – 3.600 MHz) is traditionally used for local communications out to around 30 miles or so during the day, and quite a bit farther at night. A large number of local and regional *evening* nets meet on 80 meters, often to exchange traffic via *relay* to other regional nets. Slow speed code nets can also be found within the Technician segment of 80 meters. Your very first HF CW contacts will quite likely occur on 80 meters.

During the daytime, 80 meter propagation is either by *direct wave* (line-of-sight), or by ground wave *if* you erect an antenna designed to take advantage of this mode. During the daytime, D-layer ionospheric absorption is very high, or even total, which eliminates most communication beyond the *radio horizon* which is about 30% beyond the *visual horizon.*

A typical Technician class 80 meter antenna will be relatively low dipole or *inverted V*. Such an antenna produces little if any ground wave, but can be a very effective *near vertical incident skywave* (*NVIS*) antenna at night. More on NVIS later.

To take advantage of 80 meter ground wave propagation, you *must* erect a vertical antenna (which can be a major challenge considering the long wavelength), and preferably *some kind* of ground system. The latter requirement depends greatly on your particular local geography and geology. Many hams, however, have considerable success with shortened vertical antennas, using various means of *loading*. Mobile operation on 80 meters is a popular activity, using highly abbreviated vertical whips and a basically nonexistent ground. It's safe to say, however, that nearly all 80 meter daytime mobile communication is by means of *direct wave,* not ground wave.

It's an entirely different world at night, however. As night falls, the D layer (at around 60 kilometers in altitude) disappears. You can now take advantage of F2 layer propagation (that layer is on the order of 250 kilometers in altitude), which can result in *very* long distance propagation. But before we go to *very* long distance propagation, let's talk about *somewhat* long distance propagation, via NVIS. During the evening, when the D layer absorption is gone, reliable communications on 80 meters are available out to about 100 miles radius, using simple dipole antennas. A low dipole transmits a good amount of its energy nearly straight up, which is then reflected nearly straight downward with very little loss. This is the primary mode that makes regional evening traffic nets so effective.

While 80 meters was never traditionally considered a DX (long distance) band, worldwide communications *are* possible, but you have to know what you're doing. Of course, this only happens at night, when the D layer is gone. But just because you don't have a D layer, you can't assume you'll have good DX propagation. Probably no band has more variation in DX capabilities over a given period of time than 80 meters. It's hard to spell out any rules for when 80 meter DX is going to be possible. However, here are some things you will want to know.

First of all, an effective DX antenna for 80 meters will need to be a low angle radiator. A horizontal antenna needs to have an appreciable height to radiate at low angles, over typical ground conditions. It takes some dedication to do this, of course. On the other hand, even a very inefficient vertical antenna is capable of radiating low angle signals, and this includes your typical mobile 80 meter installation. When it comes to 80 meters, a *little* power going where you want it to is worth more than a *whole lot of power* going where you don't want it to!

40 Meters: The Default DX Band

Your Technician 40 meter (7.025 – 7.125 MHz) privileges alone are worth your having a ham radio license. Numerous surveys have asked seasoned hams (the human kind, not the barnyard kind), "If you had only one Amateur Radio band, which one would you choose?" Hands down, 40 meters always wins. The reason is clear: 40 meters has more hours, days, and even *years* overall, where long distance propagation is available to *some place*. Even during the "dead" times of the 11 year sunspot cycle, worldwide 40 meter DX is possible. This is the reason so many high power shortwave broadcast stations are found on 40 meters (much to the annoyance of radio amateurs through the ages). Despite the interference the broadcasters cause in some segments of the band. hams still love 40 meters (again, for the same reason the broadcasters do). Fortunately for the Technician class operator, all the broadcast interference is on the *top end* of 40 meters, *far away* from the Technician allocation. In this regard, Techs may actually have an advantage over higher class operators!

Now, like 80 meters, 40 is primarily a night-time band. D layer absorption is high during the day (though not as high as on 80). On the other hand, ground wave propagation is even more difficult on 40 as the losses are extremely high. Nevertheless, 40 meter direct wave propagation is usually very good. For this reason, local 40 meter mobile nets were once nearly universal throughout the country, only diminishing with the advent of VHF FM repeaters, where antenna installations were far more compact.

Unlike 80 meters, it's quite practical for the average ham to build a 40 meter dipole at a reasonable height for low-angle radiation, generally up around 35 feet or so. "Back in the day," most Novices worked their first real DX on 40 meters. Most likely, as a Tech, you will too.

Skip distances tend to be very long on 40 meters, compared to 80 meters, as the signals penetrate the F2 layer farther until being reflected (actually refracted) back to Earth. Because of the long skip distances, you will usually have large "skip zones" where the signal overshoots a large swath of real estate while it lingers high in the ionosphere. This makes 40 great for very local and very distant communications…but somewhat poor in between.

Unless you use NVIS! The 40 meter band is actually pretty good for NVIS *most* of the time. However, in many cases, 40 meters is *above* the critical frequency….that is the frequency at which vertical incident signals completely *pass through* the F2 layer into outer space. This is part of the reason 40 meters can appear to have sudden "blackouts." In reality, the critical frequency has simply fallen below 7 MHz — usually right during the middle of a traffic net!

15 Meters: Daytime Delight

The 15 meter band (21.025 – 21.1 MHz) is a daytime DX band. It is sometimes called the "manic band" because when propagation is good, it's *really good,* and when it's bad, *it's really bad.* The 15 meter band can remain dead for several years of the sunspot cycle, only to scream to life during the solar peak. When conditions are good, worldwide communication is possible with extremely low power.

An unusual feature is that 15 meters is on the *third harmonic* (exactly three times the frequency) of 40 meters. This means that a simple dipole for 40 meters will work beautifully on 15 meters as well. This allows the simplest two-band operation possible; with one antenna you get a nighttime band and a daytime band.

Unlike the "low bands" you can expect essentially *zero* ground wave propagation, no matter what kind of antenna you use. You will seldom if ever hear any local network operation on 15 meters. It's really sort of an "all or nothing" band, where "all" means "DX." Contacts with closer stations are sometimes possible via sporadic E or backscatter.

Because of the short wavelength, a ham with limited space can build an antenna with appreciable gain for 15 meters. A three element Yagi, or two element quad for 15 meters can be built that doesn't blot out a large region of sky…or drain the bank account. It is a great band for experimenting with antennas. We highly recommend *Cubical Quad Notes* by the late L.B. Cebik (W4RNL, SK) for practical, low cost antenna ideas. See **www.antennex.com**.

10 Meters: Easy HF with Elbow Room

The 10 meter band is by far the largest chunk of HF spectrum available to the Technician, a full 500 kHz wide (28.0 – 28.5 MHz). Not only is there a lot of elbow room for you and your fellow Technicians, but you can also use voice and all the nearly countless HF digital modes on parts of this band too. Effective antennas for 10 meters are very small, and there's little excuse for any ham not to be able to get up and running on 10.

Like 15 meters, 10 meters is somewhat "manic." However, there is a lot more local activity on 10 during the solar "down times" than on 15. Part of the reason for this is that 10 meter local mobile activity is available day or night. An 8 foot mobile whip is a very efficient antenna for 10 meters, and 10 meter mobile is actually a lot like 2 meter simplex on steroids. The "line of sight" is far greater on 10 meters than on VHF.

Again, like 15, when 10 is "hot" it's *really* hot. Worldwide communications can be had with just a couple of watts of power, a mere fraction of the 200 W your Tech license allows. You can experiment with voice, data, CW, and have some real fun.

One of the long-standing presences on 10 meters is 10-10 International, which was formed to encourage 10 meter activity back in the 1960s when 10 meters was going through some solar doldrums. The goal of 10-10 is to work 10 other members of 10-10 International, after which you get a lifetime "10-10 number."

For the experimental minded (and cheap!) ham, it's hard to beat 10 meters. Many projects over the years involved converting CB radios to 10 meters. It's a great way to learn how to dig into a radio… without the danger of destroying an otherwise expensive rig!

Because of the fickleness of 10 meter propagation, a large network of 10 meter *beacons* has been assembled over the years. If there's any propagation at all, you should be able to tune in a beacon station, most of which are within your Tech allocation. Of course, you don't need a license to listen to beacon stations. Listening to the beckoning beacons is a great way to learn some radio science in the process, too.

6 Meters: The Magic Band

Although 6 meters (50 – 54 MHz) is VHF rather than HF, it is often overlooked in the Tech literature. A lot of new HF transceivers include 6 meters, so it's not as difficult as it once was to try this interesting band. It has many of the characteristics of both HF and VHF. Like VHF, the antennas are small, so it's pretty easy to set up an effective station just about anywhere. The band is also fairly wide, so there's not too much crowding (except when things get really hopping!). There are a number of FM repeaters on 6 meters around the country, as well. But the real action, as in the case of HF, is in the "weak signal" work...meaning SSB, CW and digital modes.

Compared to VHF and UHF, it's a bit easier to obtain good 6 meter weak signal performance without being a rocket scientist. Building low noise front ends for higher VHF or UHF bands can be a challenge for the new ham. Six meters uses fairly common technology, and transmission line losses are reasonably low. Like the higher HF bands, 6 meters lends itself to antenna experimentation.

You can experience a number of very interesting propagation modes on 6 meters. It can work in the same manner as HF, using F2 refraction. But it is also responsive to the lower F1 layer, as well as sporadic E. It can also use various tropospheric modes such as ducting. You can also do exotic things like meteor scatter on 6 meters. It is also, more than any other band, capable of using *multiple* propagation modes between distant locations. This is part of the "magic" of the magic band; you often don't know which mode is actually in play.

Since the band is so unpredictable, the secret of success is just *being there.* Successful 6 meter operators are on the air a lot. Spectacular openings can be fleeting.

Final Notes

As a Technician Class Licensee, you've *earned* your privileges. You should use them! If you avoid HF you are missing out on at least half the ham radio fun. Familiarize yourself with the *entire* Amateur Radio allocations and see if there's something that might interest you, off the beaten path. This will keep ham radio forever fascinating!

About the ARRL

The seed for Amateur Radio was planted in the 1890s, when Guglielmo Marconi began his experiments in wireless telegraphy. Soon he was joined by dozens, then hundreds, of others who were enthusiastic about sending and receiving messages through the air—some with a commercial interest, but others solely out of a love for this new communications medium. The United States government began licensing Amateur Radio operators in 1912.

By 1914, there were thousands of Amateur Radio operators—hams—in the United States. Hiram Percy Maxim, a leading Hartford, Connecticut inventor and industrialist, saw the need for an organization to band together this fledgling group of radio experimenters. In May 1914 he founded the American Radio Relay League (ARRL) to meet that need.

Today ARRL, with approximately 166,000 members, is the largest organization of radio amateurs in the United States. The ARRL is a not-for-profit organization that:

- promotes interest in Amateur Radio communications and experimentation
- represents US radio amateurs in legislative matters, and
- maintains fraternalism and a high standard of conduct among Amateur Radio operators.

At ARRL headquarters in the Hartford suburb of Newington, the staff helps serve the needs of members. ARRL is also International Secretariat for the International Amateur Radio Union, which is made up of similar societies in 150 countries around the world.

ARRL publishes the monthly journal *QST* and an interactive digital version of *QST*, as well as newsletters and many publications covering all aspects of Amateur Radio. Its headquarters station, W1AW, transmits bulletins of interest to radio amateurs and Morse code practice sessions. The ARRL also coordinates an extensive field organization, which includes volunteers who provide technical information and other support services for radio amateurs as well as communications for public-service activities. In addition, ARRL represents US amateurs with the Federal Communications Commission and other government agencies in the US and abroad.

Membership in ARRL means much more than receiving *QST* each month. In addition to the services already described, ARRL offers membership services on a personal level, such as the Technical Information Service—where members can get answers by phone, email or the ARRL website, to all their technical and operating questions.

Full ARRL membership (available only to licensed radio amateurs) gives you a voice in how the affairs of the organization are governed. ARRL policy is set by a Board of Directors (one from each of 15 Divisions). Each year, one-third of the ARRL Board of Directors stands for election by the full members they represent. The day-to-day operation of ARRL HQ is managed by an Executive Vice President and his staff.

No matter what aspect of Amateur Radio attracts you, ARRL membership is relevant and important. There would be no Amateur Radio as we know it today were it not for the ARRL. We would be happy to welcome you as a member! (An Amateur Radio license is not required for Associate Membership.) For more information about ARRL and answers to any questions you may have about Amateur Radio, write or call:

ARRL—the national association for Amateur Radio®
225 Main Street

Newington CT 06111-1494
Voice: 860-594-0200
Fax: 860-594-0259
E-mail: **hq@arrl.org**
Internet: **www.arrl.org**

Prospective new amateurs call (toll-free):
800-32-NEW HAM (800-326-3942)
You can also contact us via e-mail at **newham@arrl.org**
or check out the ARRL website at **www.arrl.org**

Matters About Matter

Some Ado about Nothing

Before fully describing how radio waves behave in empty space, it might be helpful to define what empty space is. The obvious answer, of course, is that it's space with nothing in it. However, physicists have debated for ages about just what is meant by "nothing." The most recent understanding of empty space is that it's actually a dense soup of countless trillions of particles and antiparticles continuously being created and annihilated. This probably doesn't sound very empty…or helpful. The good news is that all these particles being created and annihilated don't really matter to us much as radio amateurs. For our purposes, we can consider empty space truly empty. So called "classical" electromagnetism describes how we can have electromagnetic waves without the need for "anything" for them to propagate through. James Clerk Maxwell created a famous set of equations that show that an "ether" is not necessary for the propagation of electromagnetic waves (although nothing in Maxwell's Equations *disproves* the existence of an ether). It's just not necessary in order to make radio happen.

First Things First

In most texts on radio propagation, it's assumed that the audience has a working knowledge of electrical principles in normal conductors, such as Ohm's Law, capacitive and inductive reactance, and so forth. Although it's certainly helpful to have this information in advance, it's not strictly necessary. In fact, many misconceptions and some confusion about normal electrical circuits can be avoided by looking at the sorts of forces common to radio and regular electricity. We will find an amazing and highly comforting *reciprocity* between circuits and waves. In either case, we will start with the behavior of some fundamental building blocks of matter, primarily electrons and protons and their associated *charges.*

In addition to ignoring the "particle soup" in the paragraphs above, we can also ignore the more common *neutrons*. For electronics and radio folks, neutrons are pretty much mere intellectual placeholders; their interactions with our radio waves are essentially nonexistent.

A Weighty Matter

If you were to pick up a piece of copper wire, you would recognize it as having some *mass.* This is not the same thing as *weight.* Weight is the effect that gravity has on mass. Weight is a force, mass is not. In outer space, your piece of copper wire would have no weight, but it would still have mass. This is not merely a matter of semantics; later on we will have to have a firm grasp of *mass* to make sense of this whole radio business. Now, not only do you recognize that your copper wire has mass, but you probably also recognize it as being *copper.* What makes copper uniquely copper, and not, say, iodine? Copper is made of copper *atoms.* Each copper atom has 39 protons (and 39 electrons). You can clip the wire into as small a morsel as you like, and each morsel will be copper. This was the historic understanding of *atom*: an indivisible sample of some substance. Now, we know there are things smaller than atoms. In fact we know there are things smaller than protons and neutrons, which are made of yet smaller things called quarks and such. We don't need to worry much about quarks either, but they're fun to know about.

A proton has about 2000 times as much mass as an electron. And most types of atoms have *numerous* protons in them. So it goes without saying that as far as our *senses* go, the protons are much more important. We could probably not feel the extra weight the electrons would contribute to the piece of copper we're holding. In fact, electrons weren't even recognized until a little over a hundred years ago, whereas even the ancient Egyptians did things like *electroplating*, which is the result of movement of atoms through a chemical solution. More precisely, it is the movement of *ions*, which are charged atoms that do the work in such processes. Much more on ions later.

Not only is an electron a lot lighter than a proton, but it is a whole lot smaller. In fact, physicists tell us that the electron has no size at all…it is infinitely small. However, it does have a much larger "personal space" which makes its apparent size a bit larger. But it's still really, really small.

Charge!

As important as electrons and protons are, when it comes to radio, it's not the particles themselves that are so magical, but rather their associated *charges.* In fact, for our purposes, the protons and electrons are merely "necessary evil" *charge carriers*. Charges cannot exist without the particles, but it's really the charge we want. But before we go too far down that road, we really need to talk about what charge is. Most electronics folks have a somewhat intuitive idea of what charge is, but it is often misinterpreted.

To understand charge, we'll use a useful, but oversimplified *model.* A model is really a sort of allegory or parable to describe how something behaves. It can be extremely useful even if it's quite incomplete. So the following allegory can be used to understand charge.

Imagine a proton being the Titanic, and an electron being one of the lifeboats. Let's drop one of the lifeboats into the ocean next to the Titanic. Now, let's bring in a 30 mile per hour wind. Neither the lifeboat nor the Titanic mother ship has sails, so the wind isn't going to do much for either one of them. Let's now put a one square foot sail on top of the Titanic, and a similar one square foot sail on the lifeboat. The 30 mile-per-hour wind now has something to blow on. The sail in each case is equivalent to the *charge*. A 1 square foot sail has a charge of 1. We do need to add a small refinement to complete the picture though. If you have any maritime experience, you probably know that a sailboat can go *upwind*, depending on the *set* of the sail. There are upwind sails and downwind sails. So even though our two boats have equal size sails, they have different directions. The downwind sail on the lifeboat has a –1 charge, while the upwind sail of the Titanic has a +1 charge. For a given wind speed, each boat will have an equal but opposite force on the boat it's on.

An individual proton or electron has a charge magnitude of precisely 1; it can be no other value.

Now, it should be apparent that the Titanic with a 1 square foot sail isn't going to move anywhere near as fast as the lifeboat for a given wind speed, because of the difference in *masses.* (It will also move in the opposite direction). This is the reason why the *electron* is the more important *charge carrier* for radio generation.

As lame as this allegory is, it really is pretty good at showing electron behavior when we start moving them around with our "electrical wind," or electromotive force.

Forces and Fields

We can know all we need to know about wind without knowing what it's made of. We know it has a force and it has a direction. Although we also know (if we've been born in the last few centuries) that air is made of molecules of nitrogen, oxygen, and a few other ingredients, that really tells us very little about how it behaves. It's more important to know what it does than what it *is.*

This is precisely the case when looking at this mysterious electrical wind known as electromotive force or EMF. We don't really know what it is, but we know precisely how *charges* respond to it. A charge responds precisely and predictably to an applied EMF. To make things a little more interesting, the charge itself produces a *field*.

Chickens and Eggs

One of the more disturbing aspects of this whole force/charge business is that charges can create EMF, while EMF can move charges around. So, who's really the mover and who's the "movee"? It doesn't need to be too brain-tangling however, if we take a very mechanical approach to this. In fact, mechanical analogies work surprisingly well for a number of invisible electrical phenomena.

Imagine, if you will, an old Colonial mill stream. People in American Colonial days did a lot of things with water power. Factories of all kinds, from grain mills to sawmills, used the power of moving water to do what needed to be done, and were built near natural streams. Now imagine a paddle wheel constructed over the stream so that its blades dipped into the flowing water. The most primitive arrangement used an "undershot" wheel, which didn't involve much modification of the stream itself. More advanced "overshot" wheels were used in conjunction with man-made channels which directed the *entire* stream onto a paddle wheel). For reasons we'll see later, the *undershot* wheel is a better analogy for us. As interesting as it is to see a wheel spin around because of the water pressing against its blades, it's generally more useful to have something actually attached to the shaft…like a saw blade or a grinding wheel…so let's do just that.

Now, imagine you're a new apprentice mill operator, and you know nothing about how the mill works, other than the fact that if you show up in the barn you might get paid some dollars at the end of the day. So you walk into the little barn and see a grinding wheel spinning around. You might think to yourself that the shaft is the source of power that turns the grinding wheel. But then your new boss comes along and asks you what is making the shaft turn. Pleading ignorance, the boss patiently takes you outside the barn and shows you how the shaft is connected to the paddle wheel sitting out there in the middle of the stream. "Aha!" you say to yourself. "The paddle wheel is actually the source of power!" So, with your limited knowledge of machinery, you conclude that the paddle wheel is generating power and the spinning grinding wheel is using up the power. In modern-day, high-tech language, the paddle wheel is the *source* and the grinding wheel is the *sink.*

But this is really an incomplete picture. In fact, the paddle wheel itself is a *sink* for yet another *source*, which is the flowing water. The prime mover is really the moving water. But even that's not totally correct. It is *gravity¸* which is the original moving force for the water. So you see, depending on your point of view, various items in the entire *system* can be either a source or a sink of power.

Now, let's translate this entire business back to electrical work. If we imagine the flowing mill stream as being an electric *field*, we can see that it can produce movement of an item of some sort. The paddle wheel can be considered a charged particle in this regard. What makes this analogy work so nicely is that, at least in the case of the undershot wheel, the wheel uses only a small percentage of the total "field" of water. In electrical terms, *most* of an electrical field is wasted in that only a small percentage of its entire volume is moving actual charges around.

Hopefully, this earthy little example has shed some light on who's the chicken and who's the egg when it comes around to fields and forces. It's largely determined by where any component is in the pecking order.

Newtonian Mechanics and Other Delights

It's very fortunate that a great deal of what we observe when it comes to electromagnetic waves can be described by old fashioned Newtonian mechanics. Concepts such as *work*, *power, energy, acceleration,* and *mass* all have their counterparts in the invisible world of radio propagation. We can do a lot of shuffling around between *particles* and *waves* because of this wonderful symmetry. So, whenever we *can* use a mechanical analogy, we *will.* Again, we need to remember that a *model* is not perfect, but it's extremely useful.

One of the guiding principles of Newtonian mechanics and electromagnetism is: force (f) equals mass (m) times acceleration (a), expressed in the familiar equation:

$$f = ma$$

We can use any system of units for the particular terms, as long as we're consistent. Force can be in pounds or Newtons; mass can be in slugs or kilograms; and acceleration can be in miles per hour squared, or feet per second squared, and so on. (Nobody has ever satisfactorily defined a square second, though!)

The reason this equation is so important is that *all* radio waves are the result of the *acceleration* of charged particles. Particles have *mass*, they are accelerated by a *force*, and the overall result is the product of the force and the mass. As we move into antenna theory a bit, we will see that mysterious entities such as *radiation resistance* exist because we are doing *real work* on *real particles.*

It's a Complex World

It's rather unfortunate that we use terms such as "imaginary" and "complex" to describe mathematical methods that are inseparable from not only Newtonian mechanics, but radio itself. In fact, these mathematical concepts greatly simplify some basic physical behaviors that would be nearly impossible to describe by any other means. Sure, you can describe the motion of the planets from the point of view of the Earth being the center of the universe (as was once thought), but it's *much* simpler to describe planetary motions with reference to the Sun. Likewise, the imaginary and the complex number were not invented to make high school math more miserable; they were invented to, in the simplest terms possible, describe the universe as it really is. So, it's best to just swallow the live toad and learn how to work with complex numbers. Because….

Direction Matters

When applying Newtonian mechanics to charged particles and other entities of interest to radio, we have to know not only the magnitude (size) of the particle acceleration, but also the direction. In the real world, objects can be subjected to several forces at once, so we also need to resolve the net force applied to any piece of matter. See **Figure 1.1**. The direction of acceleration of any matter is always in the direction of the net force, but this doesn't necessarily mean the net force (and acceleration) is in the same direction as its current movement. In fact, an electron in orbit around a nucleus is in a constant state of acceleration — yet no energy is gained or lost because the acceleration is always at right angles to its movement. This condition is precisely and concisely described by the imaginary number.

We will, in addition, discover that just about everything in the universe comes in what's known as *orthogonal pairs*, that is values of some physical value that are at right angles to each other in either space or in time. Though these orthogonal pairs are related, it's also possible to totally isolate each component, not only mathematically, but physically. We will see this in action as we describe the oscillating orthogonal electric and magnetic pairs of a propagating radio wave. In this regard, we will find that radio waves are *complex*, but they aren't *complicated.* Stay with us for a while, because this is actually some very neat stuff!

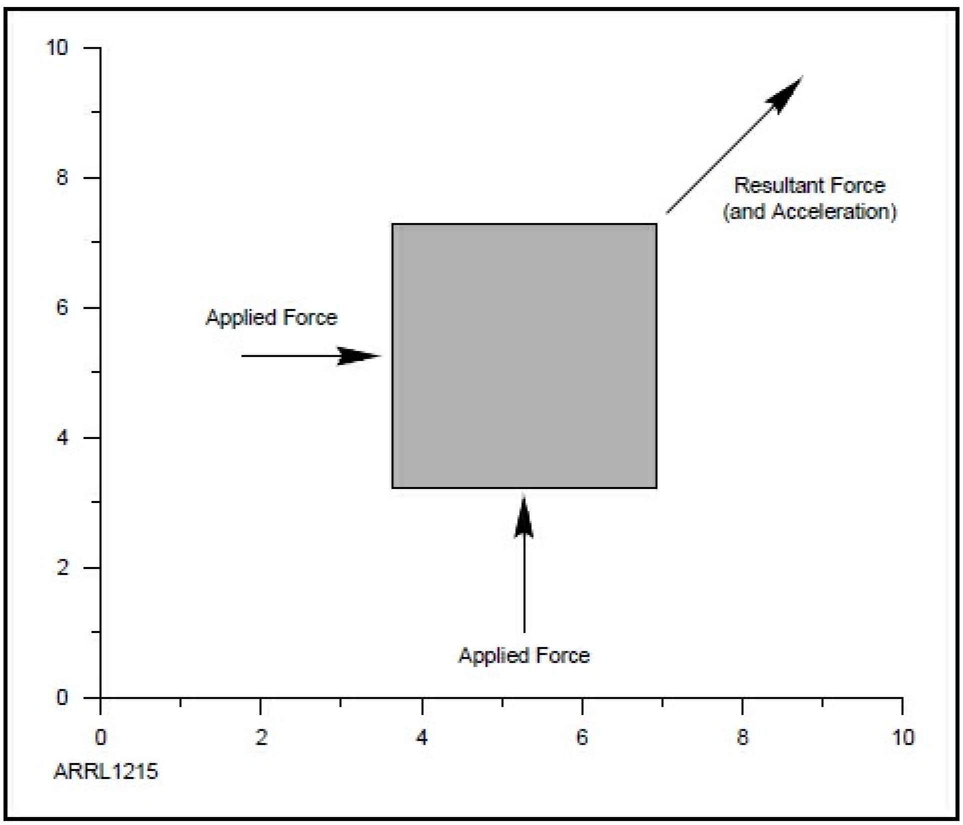


Figure 1.1 — The acceleration of a mass will always be in the direction of the *resultant* force, which may result from two or more forces. A lot of electrical principles can be explained by simplistic mechanical equivalents, as in this diagram.

Getting Real

It's probably a bit unfortunate that the square root of negative one is called an *imaginary* number. It's even more unfortunate that imaginary numbers are usually taught in high school math class with no point of reference in the real world. As we will learn in the next few pages, *imaginary* numbers have profound and very *tangible* physical significance. The significance of an imaginary number is just as "real" as a *real* number, but in a very different manner than a "true" real number. Let's forgive our mathematician friends; they had their own reasons for calling imaginary numbers imaginary, but don't let the name cause you to believe they are any less physically meaningful. But let's talk about "official" real numbers first, and things will make a lot more sense.

In electronics (and other physical sciences as well) the *real* part of a complex number signifies the transfer of *real power* from a *source* to a *load.* Another simplistic way of looking at this is that *real* components get hot! A resistor's value is *real* (resistance measured in ohms), meaning that it consumes actual *power* and converts it into either real *work* or heat. It is a very straightforward concept, and is the dominant factor in all dc circuits. In fact, you can go through life pretty nicely without imaginary numbers if all you work with is dc circuits.

Imagine That

In contrast with real values, imaginary values designate the storage of energy. Capacitive reactance and inductive reactance are designated by imaginary numbers because they describe components that don't actually consume energy (do real work), but merely store it for release at a later time. We use the word "merely" under advice, because this property is of no *mere* significance. But it *is* of a very different significance. We find that *resistance* and *reactance* are *orthogonal* values…inseparably related, but each maintaining their unique identity (**Figure 1.2**).

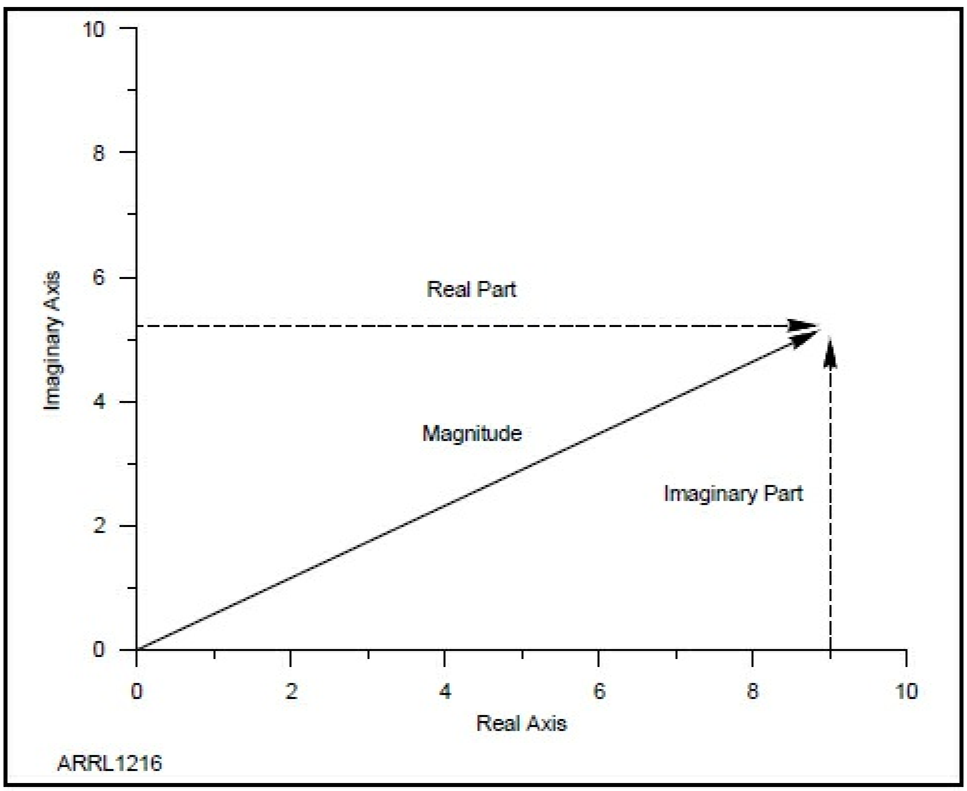


Figure 1.2 — Most physical systems in the universe can be described by a complex number. Most hams are familiar with the concept of *impedance*, which consists of the real *resistive* part and the imaginary *reactive* part. But this concept can be applied to many aspects of wave propagation as well.

Now, we do need to mention at this time that we're dealing with *real world* components in this regard. There are no perfect capacitors or inductors, and any physically "makeable" components also have some resistance. All real world components therefore are more accurately described by *complex values.* More on this when we get around to doing actual measurements!

So What

With these thoughts rattling around in our collective cranial cavities, we need to segue back into this matter of matter. A lot of what we will be dealing with as we move into propagation — especially ionospheric propagation — will be the result of orthogonal fields and forces acting on actual particles with real mass. The direction of these fields and forces relative to the particles they act upon will have profound effects in real world radio propagation.

There are a lot of alphabets in the soup, but with a little careful study we will find that this whole business of radio propagation is a lot more predictable (or at least explicable!) than commonly believed. Radio waves behave the way they do primarily because of *known* factors. The careful study of these factors will in no way subtract from the romance and mystery of radio propagation, but it will go a long way toward eliminating a lot of just plain old misinformation.

As we proceed with this fascinating study, we will keep returning to the first principles. There's a lot of material to keep track of, but as long as we periodically touch bases with the basics, things will fall into place.

The Optical Factor

When we want to look at radio propagation in its purest form, it's most helpful to refer to familiar *optical* phenomena. The properties of light were, for the most part, pretty well known a couple of centuries before the prediction and subsequent discovery of *radio waves.* You can see light directly, while a lot of the behavior of radio waves has to be determined by *inference*. We're fortunate to live in the 21st century, at a time when radio waves and light are *absolutely known* to be composed of the same "stuff," namely *electromagnetic waves.* This wasn't the case, barely a hundred years ago, when the connection between light and radio waves was tenuous at best, even among the best minds.

Although the techniques of fiddling with light differ from those used for fiddling with radio, sometimes profoundly, we're still working with the same basic ingredients. To be sure, there are different *priorities* and *significances* when comparing visible light with "normal" radio waves (see the Appendix article "Walking the Planck"), but we can still perfectly *scale* electromagnetic phenomena "from dc to daylight." What does this actually mean? It means that, while the actual *numbers* are different (vast differences in wavelength, for instance), the behavior is the same. The actual "antenna" used to emit visible light may be microscopic, or even atomic, but the fields and forces work in *precisely* the same manner as a wire dipole used for the 40 meter amateur band.

To simplify matters even further, we can use the concept of *geometrical optics* to explain many radio phenomena. By geometrical optics, we mean optics that can be explained entirely with a compass, a protractor, and a straight edge, completely ignoring any chemical or other microscopic factors. In other words we can often ignore the *materials* involved, and merely look at intuitive, familiar, and obvious behavior of light.

Mirror, Mirror in the Sky

A perfect example of geometrical optics is the common mirror. By the time most of us escape elementary school we know that for a flat mirror, the angle of incidence equals the angle of reflection (the Law of Reflection). By the time we escape high school, we probably know that this even applies for surfaces that *aren't* flat. For example, the law applies for a spherical mirror, if we use an imaginary flat surface *tangential* to the point of reflection on the sphere. See **Figure 2.1**. In fact, the spherical reflector is a pretty good *first order* approximation for the calculation of *skip distances* for certain types of over-the-horizon propagation such as sporadic-E and troposcatter, which are highly *specular* (mirror-like) reflections, which we'll describe in detail later. It's actually a pretty shabby model for *ionospheric* propagation, but we have to start somewhere. For ionospheric skip, the spherical mirror analogy is what I'd call a *half-order approximation* of skip distance — assuming we can actually estimate the height of the ionosphere — which is a fairly complex matter in itself.

Since we're going to be looking at *normal* propagation for the duration of this major section, however, it's best to look at radio phenomena which are most closely akin to optics, which implies *free space propagation in a vacuum*, and relatively distant from interfering objects. (This is one reason an antenna *test range* can be much more effective at VHF and UHF than at, say, HF or lower).

"Free space" does not necessarily mean "line-of-sight" however. A radio (or light) wave in a vacuum reflected by a perfect mirror would be considered a free space path, but not a line of sight path, obviously.

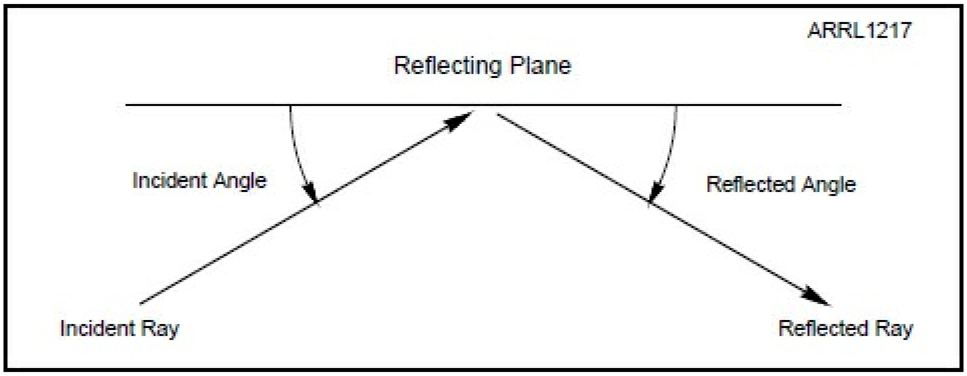


Figure 2.1 — Although true *reflection* is actually somewhat rare in ionospheric propagation, we can often simplify things by using such a rudimentary concept. The angle of incidence is equal to the angle of reflection in a simple optical reflector, such as a mirror. This also can give a good *first order* approximation for radio behavior, but don't carry the analogy too far. As Einstein once said, "Things should be as simple as possible, but no simpler."

To further define our conditions, we'd like to borrow the concept of lumped constants from the electronic circuit world. In other words, for the sake of normal propagation, we will consider a mirror, say, to exist at one discrete location, not distributed over a large volume of space. In contrast to this model, *ionospheric* propagation consists almost entirely of distributed components. The ionosphere is *not* a reflective surface like a mirror, but rather resembles a very, very long (and lossy) three-dimensional *transmission line.* This is why we will *not* include ionospheric propagation in this normal propagation section! There is nothing "normal" about ionospheric propagation, even under the most normal of circumstances!

In Prism

Of course, the mirror is not the only common optical component. The *prism* is probably the second most familiar optical device next to the mirror. A prism can come in several forms; it can be a wedge-shaped piece of glass or plastic, or it can be a collection of water droplets, which, as you have probably guessed, creates the rainbow after a shower. The prism works on the principle of *refraction* rather than *reflection*. Refraction is the *bending* of radio or light waves as a result of moving from one *medium* to another. Electromagnetic waves slow down when they encounter a new medium, and if they encounter the medium at an *angle*, the waves will be bent away from their axis. The amount of bending is a function of the *refractive index* of the medium (or more accurately, the *change* of refractive index between the two mediums).

Dispersing Some Misconceptions

We need to be careful here to distinguish between plain *refraction* and *dispersion.* The property of a prism (or rain droplets) that separates white light into the colors of the rainbow is both *refraction* and *dispersion.* Refraction is the bending of light rays through a change of medium, while dispersion is the *selective* bending of light waves with respect to their wavelengths.

Refraction and dispersion are *not* the same phenomena, but they occur so commonly together in many substances that they are often erroneously used interchangeably. You can't have dispersion without refraction, but you can have refraction without dispersion. In fact, most high quality *lenses* are specifically designed to have refraction *without* dispersion…such devices are called *achromatic* lenses. They are designed to create focused objects, not rainbows!

Likewise, in radio, both refraction and dispersion are important concepts, but they must remain distinguished from one another! Like the prism, the ionosphere is both *refractive* and *dispersive*….among a lot of other things. Again, we'll save this for later.

Let's Make This Plane

Before we get too dispersed in our thinking (pardon the pun), we should talk a little bit about *rays*. Like many other concepts in physics, the *ray* is a purely mathematical construct. There are no such things as rays…either light rays or radio rays. They are useful models to demonstrate the flow of power from a source through space, but they can be misleading without some interpretation.

Radio waves (and again, we can always replace radio with *light*), generally attenuate (get weaker) with distance, even in a perfect vacuum with no material to absorb the energy. This is a result of *geometric attenuation* which results because the "rays" (note the quotes) move away from the source in a spherical manner. If you start out with a given amount of energy, and the energy encompasses a larger sphere as it moves away from the source, the energy per square foot of sphere must decrease with distance. This is why the radiation from the Sun decreases with distance. The Sun is essentially an "isotropic radiator" that emits energy equally in all directions. (This is only strictly true in the larger sense….at great distances from the Sun. Close to the surface, the radiation of the Sun can be very directional, but that's irrelevant to this discussion).

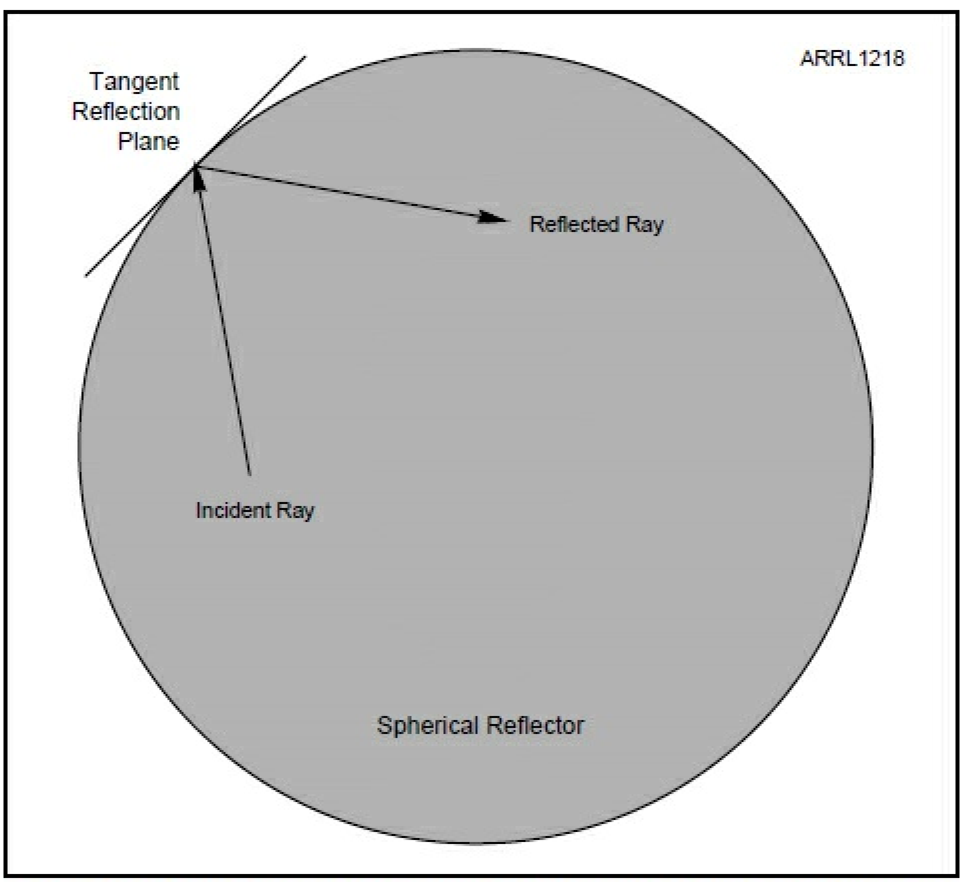


Figure 2.2 — The ionosphere can be very roughly approximated by a sphere encircling the Earth. We can use a *tangent plane* to approximate a flat reflecting surface. Again, we need to be careful not to oversimplify this process.

One notable exception to the spherical wave phenomenon is the laser, which emits coherent and *collimated* (parallel) rays that undergo little or no geometric attenuation with distance. In reality a laser consists of trillions of atom-sized antennas, each of which emits a spherical wave, but when all these are combined in proper *phase*, the result is a *planar* wave. You will not encounter true planar waves in radio — amateur or otherwise — although you can, with some effort, radiate a wave with a high degree of *collimation.* Some phased array radars can achieve high degrees of collimation…but still nothing like a laser.

So, for all practical purposes, any radio antenna, once you are several wavelengths away, may be considered a point source of spherical waves. Regardless of how much gain you start out with, the energy will attenuate at very close to the spherical formula, which follows the inverse square law. (The field strength goes down as the square of the radius of the sphere). It's convenient that this inverse square law applies to gravity and magnetism too (at least to a very close approximation), so it's easy to remember.

Fuzz and Cosmic Dust Bunnies

While geometric attenuation is a fairly universal "problem" with radio propagation, it's not the only one. The space through which we commonly send radio signals is filled with varying degrees of fuzz and cosmic dust bunnies, which cause further attenuation. We all know how fog and smoke attenuate sunlight. In a similar manner, lots of things in the atmosphere can attenuate "normal" radio waves, converting the energy into useless heat, or merely directing some of it to a less useful direction. There's a saying in physics that if you want to know what's happening, follow the heat! Now, at the risk of adding unnecessary confusion, we need to acknowledge that radio engineers and optics physicists don't agree on terminology.

Remember when we were talking about complex numbers, and how we used the *real* component to describe power being dissipated, and the *imaginary* component to describe energy storage? Well, the optics folks have it backward! We radio guys are right, and they're wrong! In passing through a lossy media (ie cosmic dust bunny-filled space) the optics folks call the *transmission* component (that is, useful, lossless, energy propagation) the *real* part and the "fuzz" loss (which actually is converted to heat) as the *imaginary* part. The imaginary component is often called the "extinction" coefficient for obvious reasons. With enough cosmic dust bunnies, the useful energy is entirely lost as heat.

The good news in all of this is that it *really* doesn't matter who's "right" as the complex math comes out with the same answers anyway, with just a sign reversal. It's really more a matter of semantics. But you should be aware that the discrepancy exists. And as we move into ionospheric propagation, we will use the optics folks' definition….the propagation constant is *real* and the dust bunny loss is *imaginary*. And, in reality, we'll discover that there is some method to the optics guys' madness.

Now Let's Be Transparent

In optics, the "dust bunny" loss is often referred to as the *opacity* of the medium. And, as you probably suspect, we have an equivalent term in radio, and we call it *absorption.* At the upper microwave frequencies, optical opacity can indeed manifest itself as radio absorption — some microwave links in Los Angeles have serious problems on particularly smoggy days. The lower you go in frequency, however, the less connection there is between optical "smog" and radio absorption.

As it turns out, at most "normal" radio frequencies, the Earth's atmosphere is almost totally transparent to radio waves. As we move even lower in frequency, however, we find that a stranger phenomenon happens. An *optically* transparent atmosphere can become highly absorptive at radio frequencies! This is particularly true at medium wave frequencies, where the so-called D-layer absorption can be very high, yet exhibits absolutely no optical effects whatsoever. Much more on this later!

Note the term "D-layer" in the previous paragraph. The word "layer" implies something akin to thin layers on an onion, but you should be aware that the ionospheric "layers" are many miles thick. That's why ARRL publications tend to refer to the D region, E region and so on, rather than the D layer or E layer.

Some Polarizing Words

In most Amateur Radio literature, the matter of wave *polarization* is thrown in as somewhat of a footnote. However, this matter is of such importance, especially in light of some more recent discoveries of radio propagation, that it merits an entire chapter of its own. In the next chapter, we'll again introduce the familiar optical properties of polarization first, and then translate them to practical radio frequency applications. We will even further explore this whole matter of polarization when we finally move into ionospheric propagation, where it becomes nearly all-important.

Polarization, Gain, and other Antenna Matters

So far, our discussion of free space wave propagation hasn't mentioned much about wave *polarization.* In a lot of Amateur Radio literature, the topic is either largely ignored, or otherwise subjected to a great deal of misinformation and tortured "explanations." In the following section, we will devote more than the normal amount of time exploring this important topic, as it is one of the areas where amateurs are most likely to be able to contribute to the state of the radio art, as well as to take advantage of the proper use of polarization for best communications effectiveness.

But First, Some Definitions

A "normal" electromagnetic wave consists of two plane waves, oriented at right angles, and electrically *in phase* as shown in **Figure 3.1**. These two *orthogonal* components are the electric component and the magnetic component (shown as vertical arrows and horizontal arrows, respectively), and their interplay results in an *electromagnetic* (*EM*) wave, logically enough. The velocity of an electromagnetic wave in free space is very close to 300,000,000 meters/second, independent of the frequency of the wave. For the sake of this preliminary discussion, we'll limit our study to waves in a vacuum, where all the relevant characteristics are consistent and predictable.

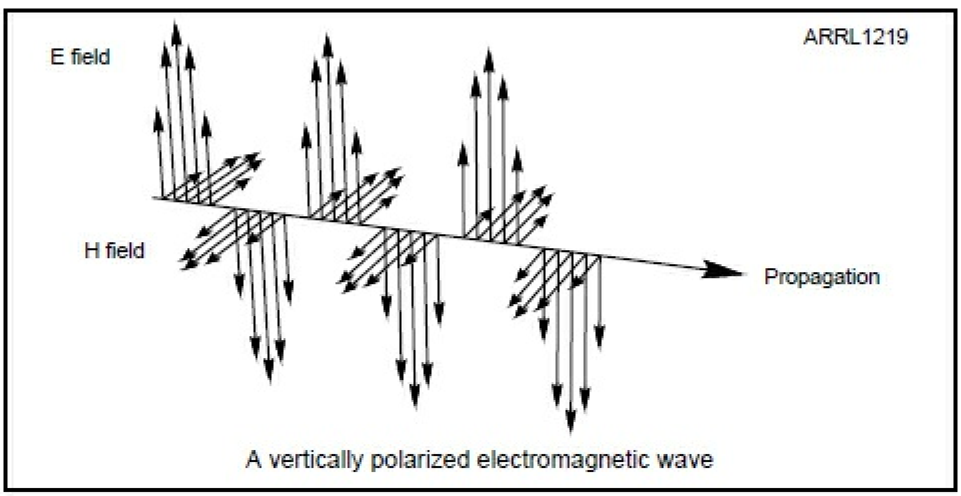


Figure 3.1 — Electromagnetic radiation consists of an electric field and a magnetic field *in phase* with respect to time, but in *quadrature* relative to position. By convention, the *polarization* of an electromagnetic wave is defined as the polarization of the electric (E) field.

With a few rare exceptions, most Amateur Radio antennas are of the *electric* type, meaning the *electric field* is the same polarization as the wire or wires used to launch the signal. A somewhat misused term, *magnetic antenna,* is used to describe certain types of compact antennas where the magnetic component is much stronger than typical, but most of these antennas are still electrical antennas by the strict definition given above.

If a wire dipole is horizontal (with respect to the Earth), the electric field of the wave it launches is horizontal. The magnetic field will be perpendicular to that wave, or vertically polarized, but by convention, in this instance the EM wave is defined as horizontally polarized as shown in **Figure 3.2**. It's somewhat more than convention, however, as we will soon see, as the *electric* field has somewhat greater practical significance for what we do as radio amateurs.

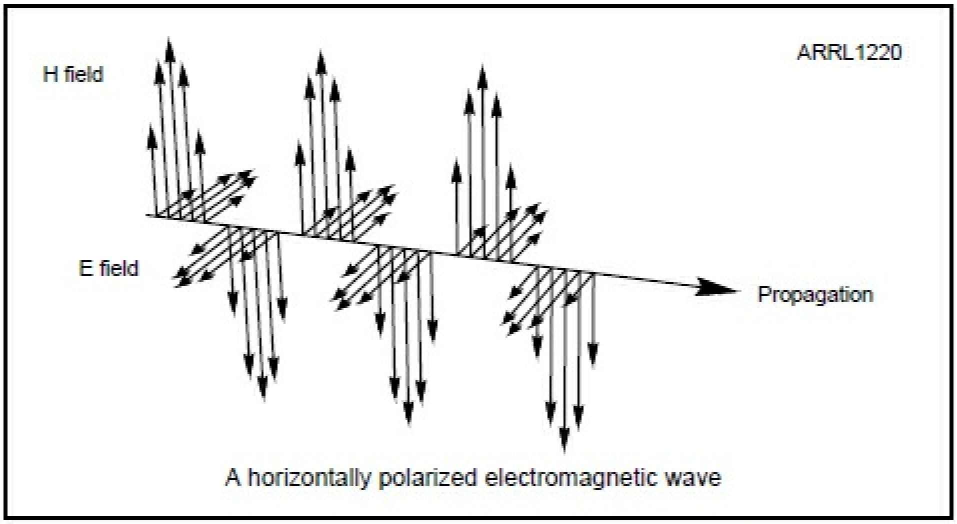


Figure 3.2 — A horizontally polarized electromagnetic wave has its electric field parallel with the surface of the Earth. In the absence of Earth (or other planet) as a reference, we may have to redefine "vertical" and "horizontal." This is especially applicable in satellite and other space communications.

Now, it is extremely important to realize that it is the *electric* field that interacts with free electrons (and not-so-free electrons). Although the magnetic field is always associated with the electric field in an electromagnetic wave, when such a wave encounters something that is *not* a vacuum, it's the electric component that does the direct interaction with matter. In fact, for most of the remaining discussion, we can ignore the magnetic wave, as it does little more than clutter up our thinking. No, we can't make it go away, but we can treat it as something akin to a mere counterbalance to our electrical wave…sort of like the heel end of a golf putter.

Being Coordinated with Your Coordinates

While it's useful and natural to refer to the polarization of a radio signal with respect to the surface of the Earth, there are some definite downfalls to this. The surface of the Earth is curved, whereas, radio waves travel in very straight lines, unless there are compelling reasons for them not to. One can very quickly lose the Earth's surface as a meaningful reference point for polarization, especially when dealing with space communications. So it is always best to use the polarization of the *transmitting* antenna as a fixed reference. This is standard practice for any kind of antenna modeling system, as well. It doesn't mean the Earth's surface isn't useful — it's just that it's often a moving target, or a very distant one.

It's also crucially important to be aware of the EM wave's *direction of travel.* As we will see later, it is impossible to fully know the polarization of a wave *without* knowing which way it's moving. In this regard, the concept of the *ray*, while, as we mentioned earlier, is only a mathematical abstraction, is extremely useful for helping us keep things straight.

It also happens to be a curious paradox that in certain cases we cannot accurately determine the direction of an EM wave's travel without knowing its polarization! Fortunately, this paradox is not too difficult to solve, at least if you have a couple of antennas. I'll talk more about this in the

"Triangulation in 3D" section in Chapter 14.

Now, recalling that it's the *electric* component that does the work for us in radio, we need to revisit the concept of *work* itself. Work is done on an electron (or any other mass) when it is *accelerated.* An electron in constant motion requires no work to keep it in motion. Nor can it impart any work to another system. (Newton's laws of motion apply 100% for any practical discussion in this entire book!) Energy is transferred (work is done) when an electric field *accelerates* an electron…and vice versa.

If we have an EM wave traveling through space, any electron it encounters will be accelerated *in the direction* of the electric component of the wave. Remember it is *electromotive force* at play. In the case of free-space propagation, the source of the electrons that will be accelerated are those contained in an electrical conductor, such as a wire dipole. Considering an infinitely thin wire, the available electrons in that wire have only one dimension of freedom of movement, that is, along the wire. Although it's a bit of an oversimplification to consider a wire as being a hose full of free electrons, for the sake of this discussion it's really not a bad model!

This means that, in order for the electric field to have any useful effect on the electrons on the wire, the wire has to be of the same *polarization* as the electric component of the incoming EM wave. When such a condition prevails, the electric component will accelerate electrons along the wire, and will do so in a sinusoidal fashion (assuming the wire is infinitely thin and a perfect conductor… obviously an impossible feat). But it's a pretty good approximation.

If the wire is perpendicular to the electric field, there is no direction for the electrons to accelerate (other than across the infinitesimal diameter of the wire). In this case, no work is done, and the EM wave passes through the wire unaffected. This is known as cross-polarization.

As we rotate an antenna from a cross-polarized state to the "properly" polarized state, we will encounter a continuously increasing amount of energy coupled from the wave to the wire. In fact, some simple trigonometry will show that the increase follows a perfectly sinusoidal pattern. Specifically, the degree of coupling is directly proportional to the cosine of the angle between the wire and the electric component of the EM wave.

This is absolutely no different from any mechanical system where we can resolve the net force on a body by looking at two (or more) different *vectors* at right angles. Imagine two people trying to push a stalled car, with one pressing on the rear bumper and one pushing against the passenger's side door at right angles. The person pressing against the door isn't doing a thing, no matter how much "work" he thinks he's doing. Only the one pushing on the bumper is causing the car to accelerate. On the other hand, if both people push on the same corner of the rear bumper at a 45 degree angle to where they want the car to go, the car will accelerate, but at a reduced rate. Only the component of their "push" in the forward direction is going to do any good.

The Tricky Part

We now come to another aspect of this matter, which can seem like a polarization issue, but it really isn't. Without understanding this, one can really "paint oneself into a corner," both in real world antenna work, and in antenna modeling.

Let's take an EM wave (we'll call it *vertical* for the sake of argument), and a vertically oriented dipole. The dipole and the EM waves have matched polarization. But now, let's lean the dipole *toward* the direction of wave travel. As you may suspect, we will now have a reduced degree of coupling between the wave and the wire. But not because of polarization mismatch. Rather it is because of a reduced *apparent length* or *projection* of the dipole. From the point of view of the EM wave, the wire appears to be *shorter* than it would if it were standing upright. In purely mathematical terms the *projection* of a line segment onto another line segment is *longest* when the two segments are parallel.

If we were to increase the "lean" of the wire toward the wave source, we would find the projection to decrease even further, until in the extreme case when the wire is lying flat, the apparent length of the wire will go to zero. This is the familiar *null* condition that one sees off the ends of a dipole. Now, you may suspect that in this case, polarization is irrelevant…or at least *indefinable*, and indeed, you would be correct.

The Polarization Paradox

We now see that there are two conditions that can result in zero transfer of energy between a wave and a wire: cross polarization and *pattern null*. How can we know which is at play, and, equally importantly, why should we care?

Let's address the first question. We showed that the relative transfer of energy between a wave and a wire is a precise *sinusoidal* function of polarization. However, the *pattern mismatch* function is not. As we tilt our wire toward (or away from) the wave source, we find the familiar "doughnut" radiation pattern of a dipole. For the geometrically minded reader, you may wonder why this is not a sinusoidal function as well; after all the geometric projection length of a wire is indeed a sinusoidal function of the tilt angle. However, we also need to consider the relative *phase angle* of the wave as it is intercepted…the part of the wire farther away from the source receives the intercepting wave a bit later. When you add these factors together, you get the familiar doughnut pattern.

So this gives us a useful hint. We can look at the *shape* of the response curve as we re-orient our antenna relative to the wave and see if it follows a sine function or not. If it does, it's a polarization issue; if it does not, it is probably a "steering" issue. Of course, we can have an antenna that is mismatched in *both* polarization and orientation to the incoming wave. And this is where things can get really interesting…both in the real world, and on paper!

Let's say we want to accurately determine the direction of arrival of a distant radio wave. Let's assume we're floating in space with nothing to adversely affect the wave itself. We don't have a clue where the signal is coming from, either in azimuth or elevation. (Just so things aren't *too* difficult, we can still see the Earth below us, so we can at least have some meaning to azimuth and elevation). For the time being, we'll assume the incoming wave is *plane polarized.* (We'll deal with circular polarized waves later on). We probably know from experience that looking for *maximum* signal strength is pretty inaccurate. Practical radio direction finders almost always use some kind of nulling method. Nulls are much sharper than lobes on nearly any practical antenna.

So, we orient our dipole horizontally, and twist it around the azimuth until we find a null. Being smart radio direction finders, we realize that a dipole actually has *two* nulls, one off each end. So we know we can be off by 180 degrees. But even so, we've eliminated about 300 degrees of direction where the signal *can't* be coming from. Checking our GPS, we find that the null is precisely East (or West).

Or is it? What if the signal is arriving from directly above (or below)? By twisting our antenna round the azimuth, we're really just changing the polarization relative to an upward or downward traveling wave! We can still have a perfect null — actually two of them — that can appear to be a pattern null. What a conundrum. What a pain. What a paradox!

Well, things aren't as hopeless as it may seem. Fortunately, we brought along a Yagi as well…not a very good one, but it's good enough. With the Yagi, we can find the *approximate* direction of arrival, removing the 180 degree ambiguity. This is known as a "sense" antenna in direction finding circles. Once we have an approximate bore sight, we can now rotate the Yagi around the axis of its boom, and look for the best signal, which means the polarization is now matched. It may take a few iterations between steering and polarization to narrow things down, but it shouldn't be too difficult. Now that we know the *approximate* direction and polarization, we can go back to our dipole, twist it to match the polarization, and then swivel it around to find the perfect pattern null. Voila!

Now, this example may seem like a rather impractical application, but as we start looking more closely at *vertical angle of arrival* of radio signals (something generally ignored by many radio amateurs) this methodology becomes much more meaningful…and necessary.

As it turns out, any "ambiguity eliminating" scheme requires either more than one antenna, or one antenna at multiple places (for example, triangulation of a distant source). You don't always have that option, however.

Back to Work

We need to always be conscious of the fact that our goal in sending out radio waves is to perform some action at a distance — in other words, to perform some real *work.* As minuscule as this work is, when we transmit a radio signal, we are wobbling some electrons at the far end (and, more often than not, some intermediate electrons along the path). There is *real power* being transferred between you and the distant antenna.

Now, the fact is that a copper wire (or aluminum tubing) is merely a convenient vessel for electrons. If there were some other way of organizing a big column of electrons without restricting them to a wire, they would be just as "wobble-able" — in fact, more so. Unfortunately, truly free electrons are an unruly lot; they repel each other with great vigor, and without some form of pipe to contain their enthusiasm, they aren't going to be of much use to us. So, despite the fact that this amazing technology is called *wireless*, we do rely heavily on wires to keep things moving along as they should.

Nevertheless, it is the acceleration of electrons that is our true bottom line. The implication is that you need electrons to interact with our radio waves, or else there would be little point to the whole exercise. Electrons will always accelerate in the direction of the electric field unless constrained by external pressure, as in the aforementioned case when a wire is at some oblique angle to the electric field doing the accelerating.

Conversely, we can always determine the radiation characteristics of an antenna by observing the direction of the electron acceleration in said antenna. This is not to say that the task is always *easy*. There are lots of electrons, and they aren't always moving in the same direction. But the *net effect* will always be the vector sum of the individual fields generated by each accelerated electron. Pretty amazing, actually.

Circular Reasoning

In this discussion of polarization, we save the best for last, namely *circular polarization.* This very special case of polarization has far more application in Amateur Radio than commonly thought. As we will learn in the sections on ionospheric propagation, nearly all HF signals are circularly polarized, although very few hams actually *use* circular polarization on HF

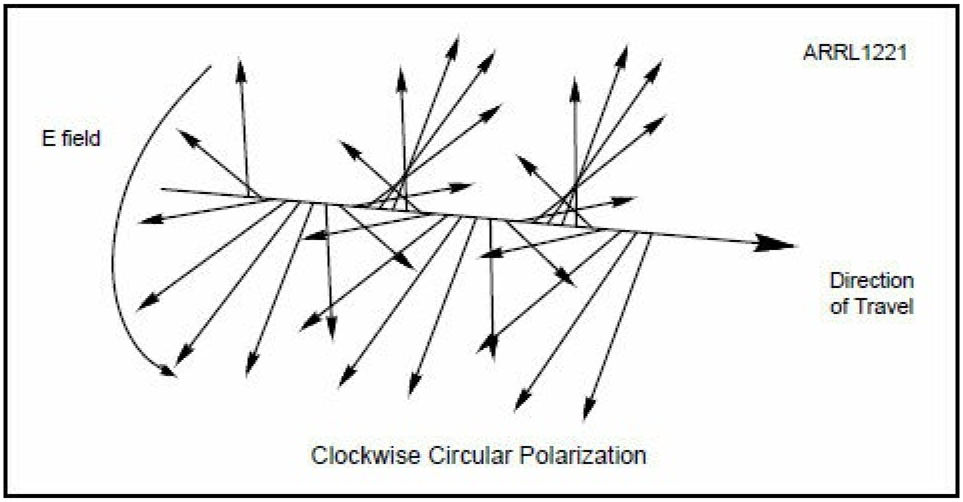


Figure 3.3 — A *circularly polarized* wave has the E and H fields in quadrature in *both* time and position. The circularly polarized (CPOL) wave is extremely important in ionospheric communications. Nearly all linearly polarized EM waves entering the ionosphere split off into two counter-rotating CPOL waves, one called the Ordinary ray, and the other being the Extraordinary ray.

A circularly polarized wave is one in which the plane of the E field (as well as the H field) rotates about the direction of transmission at precisely one revolution per wavelength of travel (see **Figure 3.3**). Think of a circularly polarized signal as "screwing its way through space." It can have either right hand or left hand threads.

The simplest way to generate a circularly polarized (CPOL) wave is to construct two dipoles at right angles and feed them 90 degrees apart electrically (in quadrature). This is certainly not the only way of creating a CPOL wave, but it is the most direct and reliable.

As it turns out, *most* real world antennas over real earth are circularly polarized *in some direction*. It may be a very narrow window where CPOL is observed, but it is there nonetheless. (Even a simple dipole over real ground has a couple of slivers of radiation that are circularly polarized, as can be clearly shown with *NEC* antenna modeling.)

As in the case of a dipole, a CPOL radio wave will transfer the maximum power when the receiving antenna is matched to its CPOL polarization. The equivalent of "cross polarization" for a CPOL antenna is circular polarization of the opposite *sense* or "handedness." A right hand CPOL antenna is cross-polarized with a left hand CPOL antenna.

On the other hand, a CPOL antenna of either sense does not know (or care) about the polarization of an incoming *plane polarized* wave. It is this property that is most used by radio amateurs, especially for satellite communications. Having a "polarization blind" antenna makes things a lot easier if you don't know what the polarization is of the incoming wave, or if that polarization happens to be slowly changing. (By the way, slow rotation of the plane of polarization, such as Faraday rotation, is absolutely *not* the same thing as circular polarization! To be circularly polarized, the rotation needs to be synchronized with the wave travel, as described above.) However, this technique only works if the incoming wave is plane polarized. If it happens to be CPOL, then you *must* match the sense of your receiving antenna to that.

Conversely, a simple dipole does not care about the sense of a CPOL wave. (It is *this* fact that accounts for the fact that very few hams know that HF signals are circularly polarized. We will discuss this at length in upcoming chapters).

Let's Reciprocate

We need to re-emphasize that all this electron wobbling is entirely reciprocal, as are all the other aspects of polarization we discussed. An EM wave accelerates electrons; an accelerated electron generates an EM wave. The direction of the electric field of an EM wave so generated is in the direction of the electron acceleration.

In free space, reciprocity is absolute and universal. However, it's important to not "over-apply" this principle to non-free-space situations. There are many other non-reciprocal factors that can occur along the path between two perfectly reciprocal antennas.

Something to Gain

This book is not specifically geared toward antenna design, as there are wonderful books such as *The ARRL Antenna Book* that go into minute detail on the matter. We defer to venerable works such as these with regard to antenna particulars. However, the antenna is the primary, (actually the only) interface we have between the "concrete" world of the electrical circuit and the somewhat less tangible realm of space. We are not reluctant to use the term *ether* to describe space. It's a perfectly acceptable description of the radio environment for our purposes.

Regardless of the gain of an antenna, once you get a few wavelengths away from the device, the actual antenna design is somewhat irrelevant to propagation. To use a crude analogy, once a golfer tees off, he doesn't have much control over the flight of the ball. All he really has control over is the initial velocity (in radio terminology, field strength) and aim. Knowledge of the environment into which the golfer is going to launch his ball is of course important. But the actual time the ball spends in contact with the club (the antenna) is minuscule relative to the remainder of the flight.

The Isotropic Bean Bag Chair

One of the concepts that seem to confuse a lot of new hams is the idea of antenna gain. Antenna gain is one of many areas where NFL reigns supreme. No, we don't mean National Football League, but rather No Free Lunch. An antenna, no matter how elaborate or large, can never radiate more total power into space than we put into it. While many radio amateurs pay lip service to this concept, by listening to a lot of on-the-air discussion of the matter, it seems that many still don't quite believe it.

An antenna can achieve gain only by taking whatever power is fed into it and redistributing it. Hopefully that redistribution results in a large percentage of the applied power going in some useful direction.

Let's imagine a beanbag chair. The skin (outer covering) of the chair is made of something like leather that is infinitely flexible, but totally non-stretchable. In other words, we can punch the bean bag chair into any shape we like, but the total surface area remains absolutely constant.

Now, let's start out by rolling the bean bag into a perfect sphere. This is an isotropic radiator… our imaginary reference antenna. The diameter of the spherical bean bag is directly proportional to the applied power. Just to pull some values out of our hat, we'll say our bean bag ball is 1 meter in radius. From our geometry classes, we know that the surface area of the bean bag, A is:

$A = 4\pi r^2$

So with a little arithmetic, we know our surface area is 12.6 square meters (m^2). And, just for the sake of argument, the input power to our antenna is 1 W. So, 1 W of power results in a surface area of 12.6 m^2. That value will never change; we're stuck with a given surface area, unless we change our transmitter power.

Now, we can squish the bean bag so a bump sticks out in some direction. But to make it stick out more in one direction, we have to squish it inward in some other direction. We can roll the whole thing into a long thin sausage if we like. This results in a lot of gain off the ends of the sausage, and a large reduction in gain in every other direction. But the surface area remains the same. (In reality, a sausage-shaped radio beam would be really hard to generate, but a skinny figure 8 is fairly simple to do.)

One thing that becomes abundantly obvious, with this entire bean bag punching business, is that no antenna can put all the energy where you want it. There will always be side bumps and dents and a certain amount of "thickness" to the beam even in the direction you want it. When we define the front of an antenna, we mean the direction in which the greatest field strength is. Or in bean bag terms, it is the direction in which the surface of the bean bag is farthest from its center.

One of the antenna types that give the cleanest pattern is the cardioid antenna. A simple cardioid will not have a great deal of gain in the forward direction but is capable of a very deep null in the reverse direction. The cardioid looks pretty much like a bean bag that's been punched in on one side with a basketball.

Now, the isotropic radiator is nowhere near as easy to create as a round bean bag. In fact, it's physically impossible. So we use the next best thing, which is the dipole antenna, which we can build. And its gain is already 2.15 dB over the isotropic radiator. To create a dipole radiation pattern take your bean bag, and with your two index fingers on opposite sides, press them together until they touch in the middle. This is pretty much your ideal dipole radiation pattern.

The RMS Field

Particularly when working with high gain antennas, there can be a contradiction between gain and efficiency. This is not necessarily a problem, but sometimes it's nice to know what your antenna efficiency is. By efficiency, we mean the percentage of power that's actually radiated, as opposed to being converted into heat. When working with a "death ray" Yagi, for instance, there is an ohmic resistance cost involved with each additional element. A large phased array of verticals also suffers from the same affliction. In most cases, a little heat loss is a small price to pay for increased effectiveness, that is, desired gain in the direction of interest. However, by calculating the RMS field, we can reveal some unexpected losses, which may or may not be remediable.

Returning to our beanbag, if we consider the skin to be the locus of all possible directions, and each of these directions has a specific field strength, we can take the RMS value of all these points and that RMS number should equal the same value, regardless of the shape the bean bag is punched into.

Of course, we'd need to take a whole lot of measurements to get the complete RMS field picture, and it might take a lifetime to do it. However, we can narrow the problem down by taking a slice through the bean bag, and taking readings that are only on that plane, say every 3 degrees or so in azimuth. In the AM broadcasting business, this is known as "running the radials" and is a required exercise to be performed every few years, or when a change is made to the array. If an AM broadcast array is shown to be less than 95% efficient, the FCC squints at a funny angle and makes all kinds of gagging sounds that you don't want to hear. Radio amateurs, of course, are not required to ascertain any level of antenna efficiency, but the exercise could be fun and instructive anyway.

A Clue

A lot of hams worry if their antenna, regardless of complexity, has lots of lobes or particularly sharp nulls. As a general rule, this indicates that the array is actually pretty efficient. Losses in an antenna system generally make the entire pattern cleaner but at a cost of overall efficiency.

A high gain parabolic dish, for example, has a lot of narrow lobes, deep nulls, and even a back lobe, to the surprise of many. The lobes can be reduced by tapering the array; that is, feeding less power toward the edges of the dish, to prevent slipover. This does indeed reduce the sidelobes and such, but at a considerable loss of gain. So, as in every other Amateur Radio adventure, there are always compromises that need to be made.

Antenna gain gives you a stronger starting point, but the rate of attenuation remains cruelly the same. Distance is the great equalizer in all antenna matters! For a little more insight into this, please read the sidebar "The Isotropic Beanbag Chair."

Coming up next, we'll talk about the process of reflection, again drawing heavily upon familiar optics.

The “Reflection” Process

Traditional Amateur Radio literature about radio propagation doesn’t deal much with molecular and atomic level processes. This is understandable; for the most part, what we experience and observe on the ham bands is the result of large scale (macro) behavior of the Earth, the atmosphere, the troposphere, the ionosphere, and so on. It’s generally the “mob behavior” of vast numbers of electrons or ions that determine the net result of a radio signal’s path from Point A to Point B. And, truthfully, you can go through a long successful ham radio career knowing little about molecules, atoms, protons, or electrons. We generally are interested in the large scale effects; after all, most of the things we like “reflecting” radio signals off of are very large compared to our radio waves and our hardware.

And yet, most of the really interesting behavior of radio propagation cannot be explained by such a low resolution approach. It is this author’s sincere belief that more rapid progress can be made in understanding and explaining radio propagation by using a microscopic approach and moving up in scale, rather than the other way around. Using this approach also requires that we do away with a lot of misconceptions, as well as a considerable history of Amateur Radio misinformation, some of it from very respectable and venerable sources, by the way.

I suspect that a good number of our readers remember the movie *Men in Black*. The mysterious agents used a device called a “Neuralizer” to remove people’s memories of encounters with UFOs. If I may, I would like to borrow the Neuralizer to remove the word “reflection” from the minds of radio amateurs everywhere, at least with regard to radio propagation. (It is certainly an extremely appropriate term when used with *transmission lines*, however. The late Walt Maxwell, W2DU’s book *Reflections* should be required reading for every radio amateur).

Not to leave a few thousand hams dangling mindlessly after being neutralized, I wish to immediately replace the word reflection with *re-radiation.* This is what is actually happening at the particulate level, and is absolutely consistent with all that we’ve learned about reciprocity. The concept applies whether we’re talking about a reflector (or director) element of a Yagi antenna, an HF skip event from the ionosphere, or the path of a radio signal to the Moon and back.

The dangers of using the term *reflection* range from the merely trivial to the egregious. For essentially free-space conditions, the use of the word reflection may give us an *incomplete* picture, but it’s functionally correct…or at least useful. On the other hand, when dealing with, say, ionospheric propagation, the term reflection is not just incomplete; it is just plain *wrong*. The processes and the *results* of ionospheric propagation are anything *but* reflective. To not understand this is to leave us confined in a purgatory of scientific error, utterly incapable of understanding (or exploiting) what’s really happening “up there.” It is this factor, primarily, that has inhibited progress in HF communications for at least the past 70 years. The good news is, this is an area where radio amateurs can still make major contributions to the state of the radio art.

The Depressing Radar Equation

For a lot of radio amateurs (and I don't exclude myself), anything that requires a lot of math can be depressing. But in this case, it's not the math that's so depressing…it's what the math clearly tells us that's depressing. To paraphrase Mark Twain, "It's not the parts I don't understand that worry me; it's the parts I *do* understand."

What we refer to here is the so-called radar equation, which basically tells us how much path loss there is when a radio signal is reflected…er…*re-radiated* from a passive, distant object (reflector or *target*).

The basic radar equation is this:

$$P_r = \frac{P_t G_t A_r \sigma F^4}{(4\pi)^2 R_t^2 R_r^2}$$

where

P_r = power received at distant station

P_t = transmitter power

G_t = gain of the transmitting antenna

A_r = effective aperture of the receiving antenna

σ_r = radar cross section of the target

F = pattern propagation factor

R_t = distance from transmitter to target

R_r = distance from target to receiver

We can simplify this equation a bit to make it more relevant to Amateur Radio than to radar, which will also bring up some of the more salient points. In fact let's remove all the variables that are *not* related to distance (such as target, or *reflector* size), transmitter power, and receive antenna gain, leaving only the path loss vs distance parameters. Notice there are two "*R* squared" factors, one for the distance from the transmitter to the reflector, and one from the reflector back to the receiver. Since the *R*s are in the denominator, we really have two "1 over *R* squared" factors. We can simplify this to $1/R^2 \times 1/R^2$, which gives us $1/R^4$.

Reinserting the left side of the equation, we have:

$P_r = 1/R^4$

Or, the power at the receiver *when re-radiated from a passive target* at mid-path is proportional to 1 over the distance from the transmitter to the target (or target to the receiver) *to the fourth power*. Yikes! Do you realize what this means? You thought the inverse square law was bad for point-to-point communications!

Figure 4.1 shows why this is so. At the left side the little black square is a transmitting antenna. Concentric circles (offset for clarity) become progressively thinner as the energy is "stretched." When they reach the reflecting surface (vertical bar at the right) the process is repeated. The bar becomes a point source, which re-radiates another set of concentric rings, each becoming thinner with distance.

We mentioned early on that, for all practical purposes, any transmitting antenna can be considered a point source, once we're a few wavelengths away. And, regardless of the antenna gain, or its original pattern, the waves form a *spherical* wave front at any appreciable distance, which means the geometrical attenuation falls off as the square of the distance. Now, any practical *target* is going to intercept a *minuscule* fraction of the total radiated power from the original source. It, too, becomes a mere point source of radiation, and thus *its* radiated signal falls off as the inverse square of distance. So we've simply multiplied these two inverse square law sources to come up with the "final answer" at the distant station. It truly is amazing that we can receive any re-radiated signal at all, when you look at these numbers.

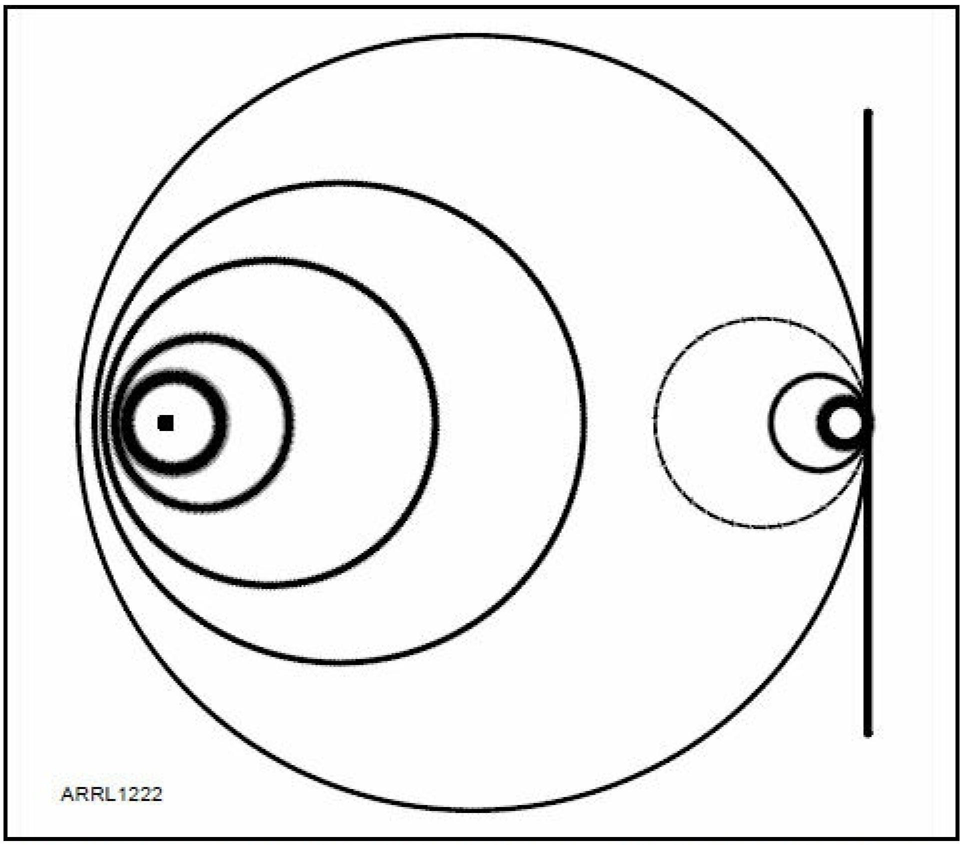


Figure 4.1 — As a radio signal emanates from a point source (small black square at the left) it attenuates as the *square of the distance.* If the waves are reflected from a distant object (vertical "wire" at the right) they are "re-radiated" as if from another point source. These reflected waves *also* attenuate as the square of the distance (R). This means that the total attenuation from a reflected object is $1/R^4$! This is the gist of the depressing "radar equation." (There are a few more constants in the equation, but the important part is the attenuation versus R). Despite the tremendous "geometric attenuation" this process undergoes, hams very effectively use reflected radio signals in a variety of situations.

But Now for Some Good News

Fortunately for us radio amateurs, the situation isn't quite as bleak as the numbers suggest. Remember in the original radar equation we had this σ symbol? This is the radar *cross section*…that is, the effective size of the target. Now, granted, a dipole 300 miles away is a pretty microscopic target. However, in most applications we're concerned with, we are attempting to bounce a signal off a considerable chunk of sky! In fact, the cross section of just about anything we normally use for beyond-the-horizon communications is pretty huge….it can be a "patch" tens or hundreds of miles across. Or in the case of "dark side" propagation, it can be a hemispherical surface thousands of miles across! So, we can breathe somewhat of a sigh of relief, knowing that re-radiated ham signals of usable strength are certainly possible. Of course, we knew that already, didn't we?

There is more to cross section than mere *size.* The reflectivity, or lack thereof, of the material making up the desired "radio mirror" can vary widely, from almost nothing, to almost perfect. Large bodies of moist air can re-radiate radio waves — but generally very poorly — so the effective cross section of even large meteorological regions can be very small. And, even if it is highly re-radiative, if it doesn't re-radiate in a useful direction, its cross section is also going to be very low. The radar cross section of a polished metal sphere is pretty much an infinitesimal point, regardless of the diameter of the sphere! The reason for this is obvious. Any signal reflected other than perfectly normal (perpendicular) to the surface of the sphere is going to be scattered off into oblivion as shown in **Figure 4.3**. The perfect stealth aircraft would be a perfectly round polished metal sphere…except that it would probably be pretty hard to fly!

At this point, you will probably wonder how moonbounce (EME) is even possible, since the Moon is a sphere. Actually, it *wouldn't* be possible, except for the fact that the Moon is very rough! There are enough nooks and crannies and crags scattered across the surface of the Moon, that, statistically speaking, it looks almost like a disk of the same diameter.

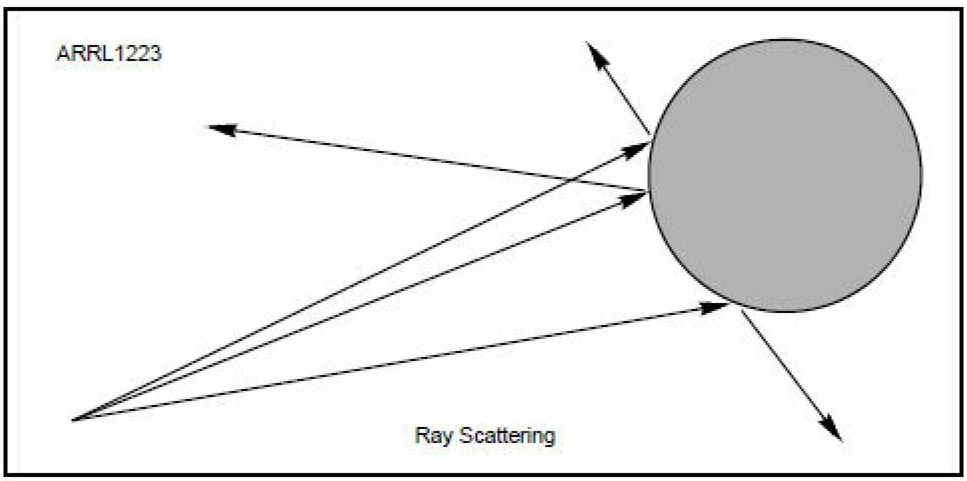


Figure 4.2 — The perfect stealth aircraft (invisible to radar) would be a polished metallic sphere (though it may be a bit difficult to fly!) A radio wave reflected off the sphere will be scattered off into space *unless* it happens to strike the sphere precisely *normal* to its surface, in which case it reflects back to the source. Thus, a perfect sphere, no matter how large, has a *radar cross section* essentially the size of a pin point. The reason we can get usable reflections off the Moon is because the Moon is very rough!

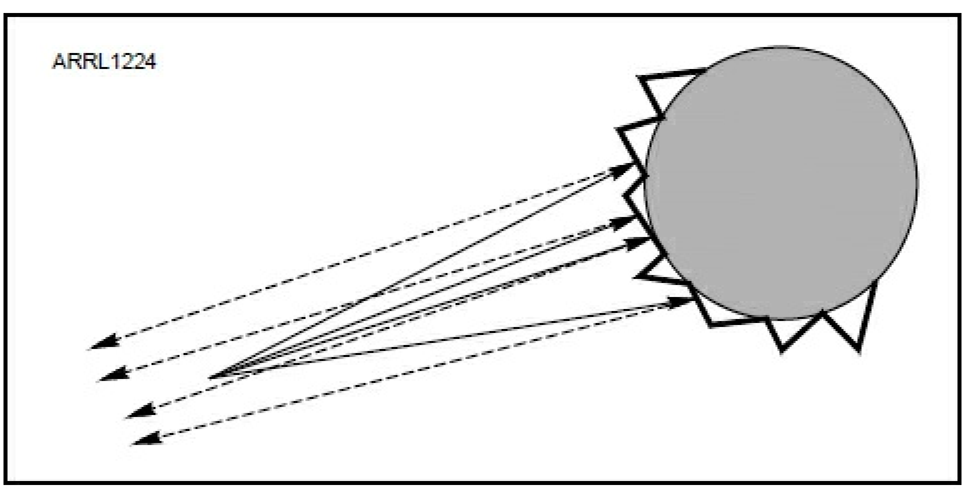


Figure 4.3 — Imperfections on the Moon's surface (shown greatly exaggerated) statistically improve the chances of at least *some* signals being reflected in the general direction of your receiver.

Tipping the Scales

When looking at the process by which electromagnetic (EM) waves accelerate electrons and, conversely how accelerating electrons create EM waves, the question may come up about wavelength and frequency. For most Amateur Radio purposes, the wavelengths involved are so huge compared to the size of electrons that for free space propagation, wavelength is irrelevant. It is only when EM waves approach the dimensions of atoms (such as in visible light) that we need to worry much about *scaling factors*…that is the amount of energy as a function of frequency. This is all neatly tied up in a figure known as Planck's constant, which is fairly irrelevant at Amateur Radio frequencies.

Any practical radio wave is going to be trillions of times larger than the dimension of an electron, so we can safely simplify a lot of things, at least when free space conditions prevail. At visible light and shorter wavelengths, on the other hand, physicists speak of the relationship between energy and frequency…we can actually ignore that. Radio waves have amplitudes and frequencies, and they are independent factors. We can have any combination of frequency and amplitude we desire, with no complicated scaling factors to contend with. For more on this, see the Appendix article "Walking the Planck."

Chorus Line

When one defines a certain material as being a good *conductor*, it's generally meant that the material has a lot of free electrons. While this is true, it's not a complete picture. Yes, we need free electrons in order to have an electric current, but those electrons need to have a somewhat unobstructed path in order to be of much use, at least as far as we're concerned.

The Ground Wave

From a historical perspective, the first beyond-the-horizon radio propagation was the result of what is known as *ground wave* propagation. In recent years, most Amateur Radio literature has relegated ground wave propagation to a mere footnote; however recent resurgence of interest in medium wave (300 kHz to 3 MHz) and lower frequency propagation merits a closer look at this classical means of long distance propagation. We can even learn a bit about ionospheric propagation by looking at ground wave propagation in detail. Since most radio amateurs live on the Earth, it's probably a good idea to *talk about the Earth itself, and how it affects radio propagation.*

Down and Dirty

Now that we know a bit about how radio waves travel through free space, let us now look at another very common *medium* that we'll encounter in our day-to-day Amateur Radio operation. We're talking about dirt, the stuff that covers most of the Earth that we live on. (Of course, most of the Earth's surface is *water*, but relatively few radio amateurs live on the water, or operate from that environment. However, we will dedicate an appropriate amount of discussion to *sea* propagation toward the end of this chapter).

Any medium other than a vacuum has two fundamental properties: electrical resistivity and dielectric constant. Both of these have important, but *very different* effects on radio waves that pass through a medium. And, by now, you should not be surprised to learn that *dirt*, or any other non-vacuum medium has *orthogonal* properties; in other words, it is described by a *complex number.* Without taking into account the fact that *dirt* has a complex impedance, we can never fully explain (or exploit) ground wave propagation. Furthermore, nearly any practical amateur antenna is going to be affected by the proximate ground, regardless of whether we're actually *using* ground wave propagation or not. So, bear with us as we take what may seem to be an inordinate amount of time discussing the nature of dirt.

Intrinsic and Extended Properties

Any radio amateur should be familiar with the concept of *resistance.* Resistance is the opposition to the flow of current in a particular conductor or other component. It is measured in *ohms*, of course. A less familiar term to many radio amateurs is *resistivity* which is an intrinsic property of the material in question, say copper. There are actually a few forms of resistivity: *linear resistivity*, *surface resistivity*, and *volume resistivity*. Linear resistivity is probably the most intuitive; it is measured in *ohms per inch,* or *ohms per foot.* A certain gauge of copper wire, for instance, has a certain linear resistivity. However, even linear resistivity is a somewhat *derived* value, as there aren't too many natural objects that come in the form of a wire!

A more fundamental value, something that is more *intrinsic* to the material in question, is *volume resistivity.* If you take a cube of a certain volume of copper, and measure the resistance from one face to the opposite face, say a meter away, you will have a resistivity measurement in ohm-meters, or ohm-inches, or some such. If you have a cube of material with a conductive plate on a pair of opposite surfaces, the substance under question can be considered to be forms of a whole lot of parallel resistors (**Figure 5.1**). The total resistance is inversely proportional to the cross sectional area of the sample, and proportional to its length.

The situation is quite a bit more complicated if you want to measure ground resistivity (or its reciprocal, ground conductivity) by means of a couple of stakes driven into the ground, the standard means of determining local ground conductivity. The reason is that you have a whole lot of nonlinear paths between the two stakes, as seen in **Figure 5.2**. Solving this requires the use of elliptic integrals, which are not much fun for the average math student. Fortunately, there are charts available that run these calculations for you, based on just one measurement.

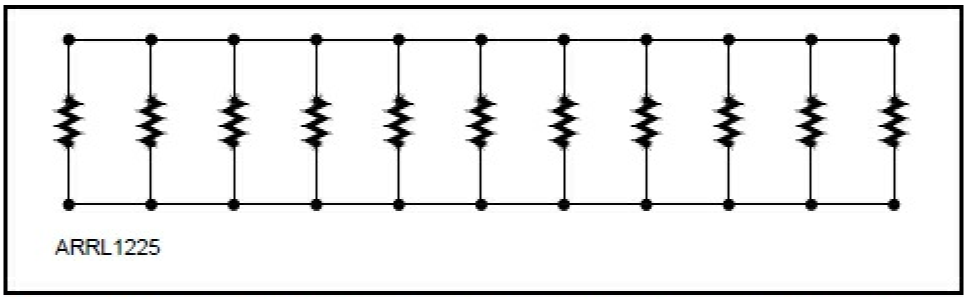


Figure 5.1 — The electrical path through a solid mass, such as the Earth, can be extremely complicated. In the simplest form it appears like a lot of resistors in parallel...but not quite. There's a lot of "lane changing" going on throughout the journey. Ground conductivity is much more easily measured than calculated.

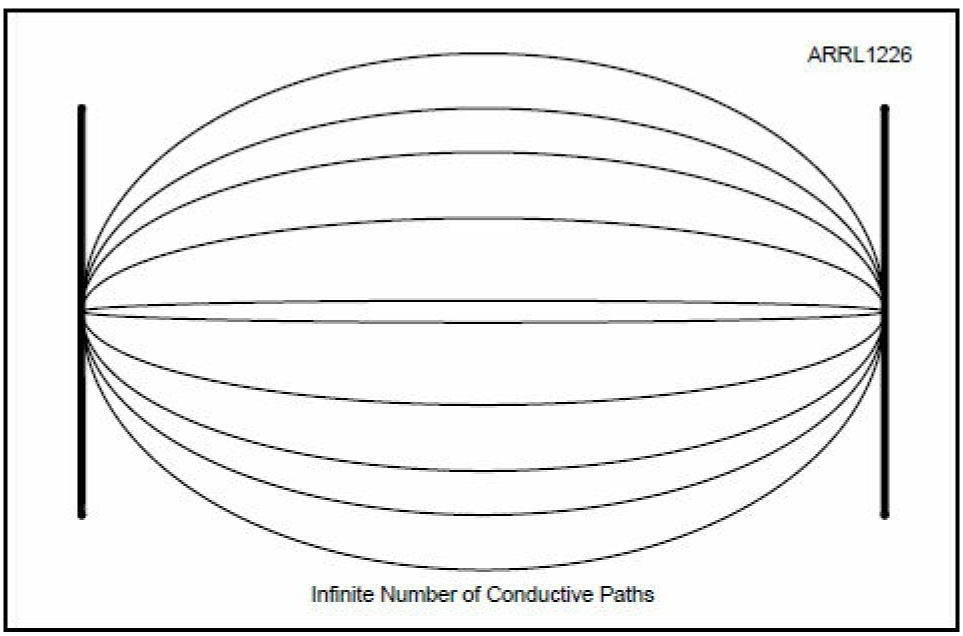


Figure 5.2 — Calculating ground conductivity can be a real challenge. If you drive two rods into the ground some distance apart, you will have an infinite number of elliptical paths for current to flow. The length of each of these paths can only be determined by extremely difficult *elliptical integral calculus*. Fortunately, we have charts and tables to help us figure these things out. Even more practically, we can do some fairly simple measurements to determine our local ground conductivity.

Now we get to the fun part. Ground conductivity (or resistivity) is only part of what we need to know about dirt. Equally important (and frankly, a whole lot more interesting) is the matter of *dielectric constant.* As we discussed Chapter 2's sections on optics, a medium with a different dielectric constant than vacuum has a different velocity of propagation than does free space.

The combination of these two properties, volume resistivity and dielectric constant, has two primary effects on any radio wave passing through. It absorbs energy, causing attenuation, and it slows the radio wave down. (In actuality, there are some media which actually have a negative dielectric constant, causing the phase velocity to be *greater* than that of free space, but we don't encounter such oddities in most of Amateur Radio.)

The attenuation is fairly straightforward, and easy to explain. The electric field of the wave, in passing through a poor conductor, causes a voltage to be impressed across the conductor, which in turn causes current to flow in the direction of the instantaneous E field. Since it's not a good conductor, there is heat loss (ohmic heating), which is real and measurable. This heat loss is relatively constant regardless of the direction of wave travel through it, assuming the material (dirt) is isotropic — that is, containing the same resistivity throughout the travel of the radio wave. This is not always the case, however, as the bulk resistivity of real world dirt can vary considerably with depth. This can be a bit tricky to measure, but we can determine some of this by *inference.*

Now, for the sake of simplicity, let us consider a radio wave coming straight down through the air and encountering typical earth perpendicular (or *normal*) to the surface. We know that, since the dirt is *partially* conductive, there has to be a certain quantity of free electrons available. Since the process of absorption and re-radiation applies, regardless of the medium, we see that electrons are accelerated to some degree at each increment of depth. If we imagine the dirt as being an extremely large number of extremely thin layers of conductor, we will have countless "reflections" occurring along the path of the descending radio wave. Each layer absorbs and reradiates the impinging radio signal. However, each layer radiates both in a *forward* direction (to the next lower layer) as well as backward (in an upward direction) This bucket brigade action of each layer absorbing and re-radiating to the next layer has the bulk effect of slowing down the radio wave propagation.

Although the internal workings of such a system can be extremely difficult to completely visualize and analyze, in general, the *effect* is that of a single sheet reflector far below the surface of the Earth. We can replace all these partial reflectors with a single sheet with an effective depth of some finite distance below the surface. As we start looking at the optimal height of an antenna above ground, we can use this effective depth to great benefit.

From a circuit oriented point of view, we can consider the attenuation or loss as the resistive component, and the *delay* as being akin to *reactance.* We don't want to over-extend this analogy too far, except to realize that both components are intertwined. And, as in optics, the total value is a complex number, one axis being resistive, and the other axis being reactive. Again, we need to defer to the optics definition, per convention, and designate the dielectric constant as the *real* component, and the loss component (also called the loss *tangent)* as being the *imaginary* component. Again, this is at odds with normal concept of *real* in the physical sense.

The important thing to remember is that ground *is* indeed a complex figure, and we need to address the dielectric constant just as much as the resistance.

Grazing Paths

Things become even more interesting when the radio signal enters the Earth at an angle, rather than normal to the surface. Let's look at the extreme case, when the radio wave is coming in *parallel* to the surface of the Earth, the low-grazing path. To make things even more extreme (and interesting), let's imagine the Earth's surface as being a perfect conductor, but all that conductor is confined to a very thin skin; in other words the Earth is made of a hollow, thin-walled copper ball. Let's further define the situation and stipulate that the wave is *horizontally polarized*, that is, the *electric component* of the wave is horizontal. The electric component, in free space, attempts to apply a *voltage* across its two "terminals." Now, obviously, a wave doesn't have terminals, *per se*. It does have two polarities, that is, the opposite amplitude extremes. If we were to apply this voltage source to a short circuit, what happens? Well, the voltage source is shorted out. It will *attempt* to cause infinite current to flow through the Earth's surface (in the same direction as the voltage difference of potential), but that can't happen. What does happen is that the difference of potential (the E component) is reduced to essentially nothing, and the wave ceases to propagate at all.

Now let's change the polarization of the wave to vertical; that is the E component is perpendicular to the Earth. We can imagine half of the E component being "underground" and half being above the surface. Even if the copper is perfectly conducting, we will only short out the "center" of the E-wave, where the voltage is at or near zero. The thin-skinned Earth will have essentially no effect on the vertically polarized radio wave. We can make a broad statement therefore, that all true ground wave signals are vertically polarized. By true ground wave, we mean radio signals that propagate for an appreciable distance in contact with the surface of the Earth. This is in sharp distinction with radio waves either bouncing off the ground, or with space wave propagation, which, while in proximity to the ground, makes no direct contact.

We can show that the surface of the Earth itself is a polarizing device. But it's actually a lot easier to demonstrate this principle with a simple optics experiment. All you need is a microscope slide, a point source of light (a light bulb or fluorescent tube works fine), and a polarized sheet of plastic (a pair of polarized sunglasses will do).

Set the microscope slide down on a table so you can see the reflection of the light bulb in the slide, at about a 45 degree angle
(**Figure 5.3**). Now, look at the reflection through the polarized lens, and rotate the lens around its plane. The reflection from the glass slide will become much dimmer at some point of the lens's rotation. At this point the polarized lens is 90 degrees with the vertically polarized reflection from the glass slide, proving that the reflected ray *is* indeed polarized. This is really a fun and astonishing demonstration. With a little bit of fiddling around you can adjust the angle of *incidence* to a point where the reflection *completely* disappears when it's cross-polarized; this angle is known as the Brewster angle. In radio, we have what's called the "pseudo-Brewster" angle, because the reflection point is nowhere nearly as well defined as it is with a microscope slide. However the underlying physics is nearly identical.

The pseudo-Brewster angle has some significance for radio signals arriving at medium vertical angles. But now, let's concentrate on very *low* angles of arrival…signals running essentially parallel with the Earth's surface.

A vertically polarized, horizontally moving wave can travel with its bottom half below the Earth's surface, and the upper half above the Earth's surface. In this state, we have two parts of the same wave traveling through different media…and at different velocities. This would theoretically require the wave to "break" at the demarcation between the two media as shown in **Figure 5.4**. However, this is a physically impossible situation. You absolutely cannot have a "split wave" propagate. Only continuous waves can propagate.

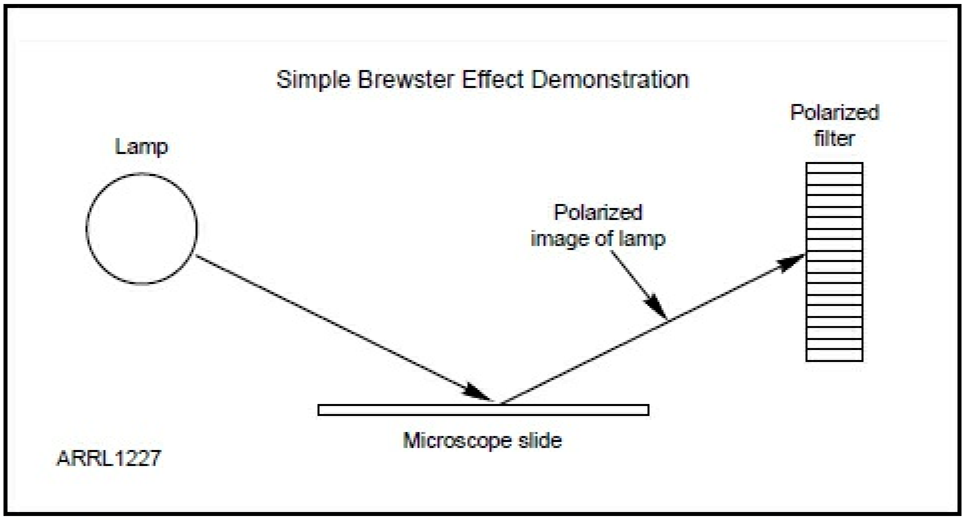


Figure 5.3 — Any "true" ground wave must be vertically polarized. In addition, any wave that *reflects* off the ground at a low angle also becomes vertically polarized. This effect can be easily demonstrated with simple optical apparatus. At lower frequencies, (radio frequencies), the so-called "pseudo-Brewster" angle applies. Although the physical principle is identical, the effect is not as well-defined in radio as it is in optics.

We *can*, however, "reconnect" the split waves by allowing the wave to bend downward. We now have a continuous wave with a slightly different velocity between the top and the bottom halves (**Figure 5.5**). It is this action that allows true ground wave propagation beyond the horizon. In fact, this is also very effective for long distance maritime communications, as saltwater has a very low loss tangent, but a fairly high dielectric constant. Again, it's important to realize that *only* vertically polarized signals can propagate in this manner. With this in mind, it's almost a given that any long distance medium and long-wave signals will be most effectively received with vertically polarized antennas.

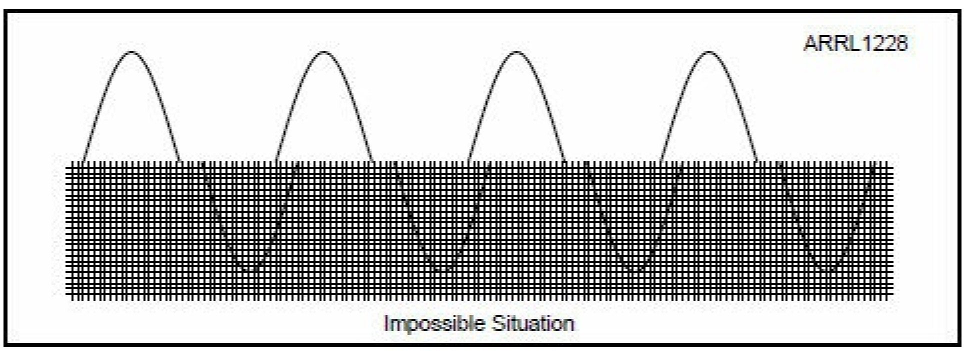


Figure 5.4 — An electromagnetic wave travels more slowly through a medium like dirt than it does through air or free space. This presents an interesting scenario when *part* of the wave passes through air and part through Earth. The two different speeds would require a "breaking" of the wave at the junction between the two media, as shown here. However, the laws of physics won't allow this to happen. All electromagnetic waves must be *continuous functions*.

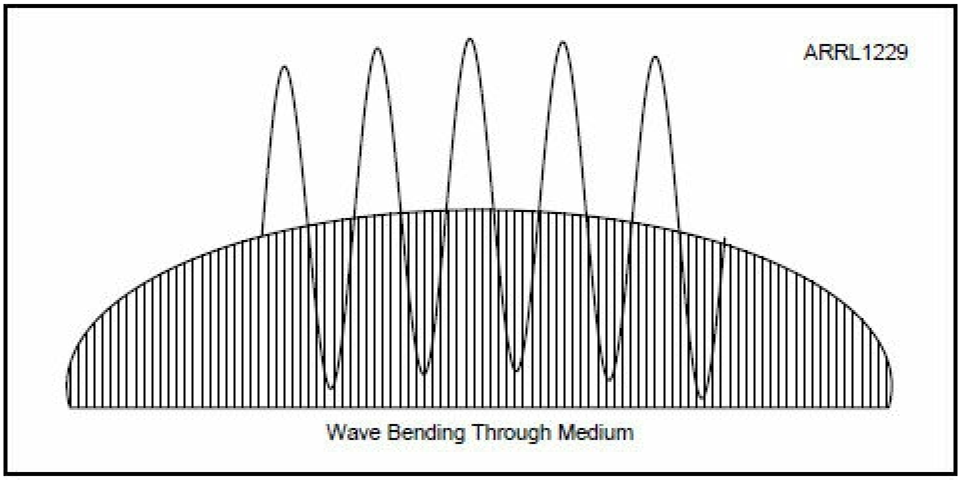


Figure 5.5 — To resolve the impossible situation in Figure 5.4, the wave must *bend* toward the slower medium to keep the wave "connected." This the process that causes *ground waves* to follow the contours of the Earth. Perhaps the "Slinky" analogy might help. If you lay out a Slinky on a table in a straight line, and compress the coils on one side (the equivalent of slower propagation), the Slinky must bend in the direction of the compressed side of the coils.

An interesting antenna that merits special consideration in this regard is the Beverage antenna, which operates in a rather different fashion than most conventional Amateur Radio antennas. A true Beverage antenna, while being physically constructed of a long horizontal wire, is actually *vertically polarized,* and very *strongly* so. Ordinarily a wave arriving "end on" to a wire induces absolutely no voltage along the wire, which is why a dipole has a perfect null at the ends. However, as we have shown, a low angle, vertically polarized wave *leans* forward slightly. This leaning action creates a small voltage differential component *along* the wire. It should be evident that a Beverage would not work at all in free space, but rather relies on the Earth's dielectric constant for its function.

One of the interesting results of the polarization filtering of the Earth's surface is that, since *only* vertically polarized signals can propagate along said surface, one may conclude that low angle incoming radio waves are inherently vertically polarized. They may or may not be. As we will discuss in great detail later, nearly all practical HF *ionospherically propagated* signals are circularly polarized. However, when such a wave encounters the Earth at a grazing angle, the horizontal component is suppressed, leaving only a vertically polarized signal as a result. Much erroneous thought about ionospherically propagated waves is a result of this phenomenon, and one has to either get a fair distance away from the Earth's surface, or use some other clever tricks to eliminate the polarization filtering of the Earth to learn the true nature of ionospheric sky waves.

Seafaring Signals

Since fully two-thirds of the Earth's surface is water, we would be remiss if we didn't discuss some of the particulars about radio propagation over the sea. As we mentioned at the beginning of this article, the first truly long distance radio signals were propagated over the world's oceans. Radio is actually very good when you're at sea, for a number of reasons. Radio amateurs who are privileged to operate on or near the ocean are always delighted at the wonderful conditions this nearly always affords.

Because sea water is highly conductive, it forms an excellent reflector for signals arriving at high angles. This can result in numerous low-loss hops across long distances, particularly noticeable at HF frequencies, but certainly not restricted to that part of the spectrum. Reflections from terrestrial Earth are generally quite lossy, as we have described above. In addition, the ocean's surface is, on the average, a lot *flatter* than land; it forms a much more effective reflecting plane for downward arriving signals. Just about any experienced DX operator can testify as to how much easier it is to operate long distances over ocean paths than over land-locked ones, at least on the average.

But you don't have to be on a ship to take advantage of this. Just being *near* the ocean, whether on an island or a large coastal area, can give you a tremendous advantage both for high angle signals and for low angle ones. Any good antenna modeling program takes into account not only the ground conditions directly below the antenna, but also on conditions many wavelengths from the antenna. Large bodies of sea water are good ground planes!

Fresh Ideas

What about *fresh water*? This is a very good question! We do have a few ocean-sized fresh water lakes in the United States, as well as other areas of the world. (The US Great Lakes are among the world's largest, however). Is there any advantage to working on or near a Great Lake…or just a great lake?

For a useful answer to this, we need to talk a little about skin depth. For a quick background, we'll refer you to an excellent article by Rudy Severns, N6LF, on skin depth of soils. Rudy's article may be found online at **www.antennasbyn6lf.com/files/ground_skin_depth_and_wavelength.pdf**

Although, as you can see, the formulas are a bit involved, the conclusions are fairly straightforward. The better the conductivity of the medium, the shallower the penetration of an incident radio wave. The skin depth (penetration depth) of average sea water to HF radio signals is about six inches! This means that there is no way you can use your HF "manpack" radio to talk to a boat on the surface while you're scuba diving down at 50 feet…or 5 feet…regardless of how much power you have! (Not to mention that building an antenna that works under sea water without shorting out is a real trick in itself!)

Because the skin depth is so low for salt water, as well as frequency dependent, we find that the only frequencies available for effective submarine communications are in the ELF (Extremely Low Frequency) region, generally a few hundred Hertz. Since most hams probably aren't going to be doing much submarine communication, this may be a moot point, but it does demonstrate some interesting physics.

Now fresh water is an entirely different animal. While the dielectric constant (propagation velocity) of sea water and fresh water are about the same, the conductivity of fresh water is nearly zero, and so the skin depth is very large. In fact, it would be no problem to communicate with submarines in the Great Lakes with HF radio methods. To the best of our knowledge there aren't a lot of submarines prowling around the Great Lakes, but one never knows. Likewise, it would be no problem to use your 2 meter handheld at the bottom of your swimming pool. From a more practical standpoint, because of the essentially non-reflective properties of fresh water, there would be minimal advantage of operating from or near a large body of fresh water. Many hams will dispute this, however; this author knows a few hams near the Great Lakes who vehemently assert that it's just as good as living near the shore of the Atlantic Ocean, propagation-wise. There are certainly a lot of other factors involved in any discussion of this nature. This may be difficult to conclusively prove one way or the other, and certainly merits some systematic investigation.

In later chapters we'll discuss more specific methods one can take advantage of one's local and not-so-local ground conditions. We trust that this introductory chapter on the subject shows that since most of us operate quite close to the Earth, the properties of the Earth itself have profound effects on most of what we do as radio amateurs.

Demystifying the Ionosphere

It is safe to say that one of the most baffling aspects of all of Amateur Radio is long distance ionospheric radio propagation. It is also one of the most fascinating aspects of our hobby. And for at least a couple of generations of radio amateurs, ionospheric propagation was central to most Amateur Radio activity as well as a focus of a good portion of on-air technical discussion, and with good reason.

Like the weather, everyone talks about propagation, but nobody does anything about it. Despite the fact that the average ham has only a skeletal concept of how the ionosphere really works, for nearly a century hams have been able to effectively exploit this often fickle entity. It might be a little audacious to suggest that we can demystify the ionosphere entirely. And, in fact, we probably wouldn't want to, as the mystery and romance of the unpredictable radio sky is a huge part of Amateur Radio's appeal.

Nevertheless, there are a lot of things about the ionosphere we *do* know, and by starting with a firm foundation, we can not only enjoy Amateur Radio to its fullest, but also explore many of the more unusual aspects of Amateur Radio. It is a subject of inexhaustible intrigue.

Models and Modes

Whenever our understanding of any scientific field is incomplete (and that pretty much includes all of them), we must resort to the use of *models.* A model doesn't have to be complete or even absolutely accurate in order to be incredibly useful. Whenever studying something as complex as the ionosphere, we need to use models based on what we can consistently observe, and then extrapolate them to (hopefully) logical conclusions.

The ionosphere is just one such entity; there are many models to describe what we observe, based on what we can reproduce under controlled conditions. As you might suspect, however, the ionosphere is *not* a controlled environment. We *think* we know what most of the contributing factors to its behavior are, but there's still a lot to learn. And that means there's still a lot of work to be done by radio amateurs.

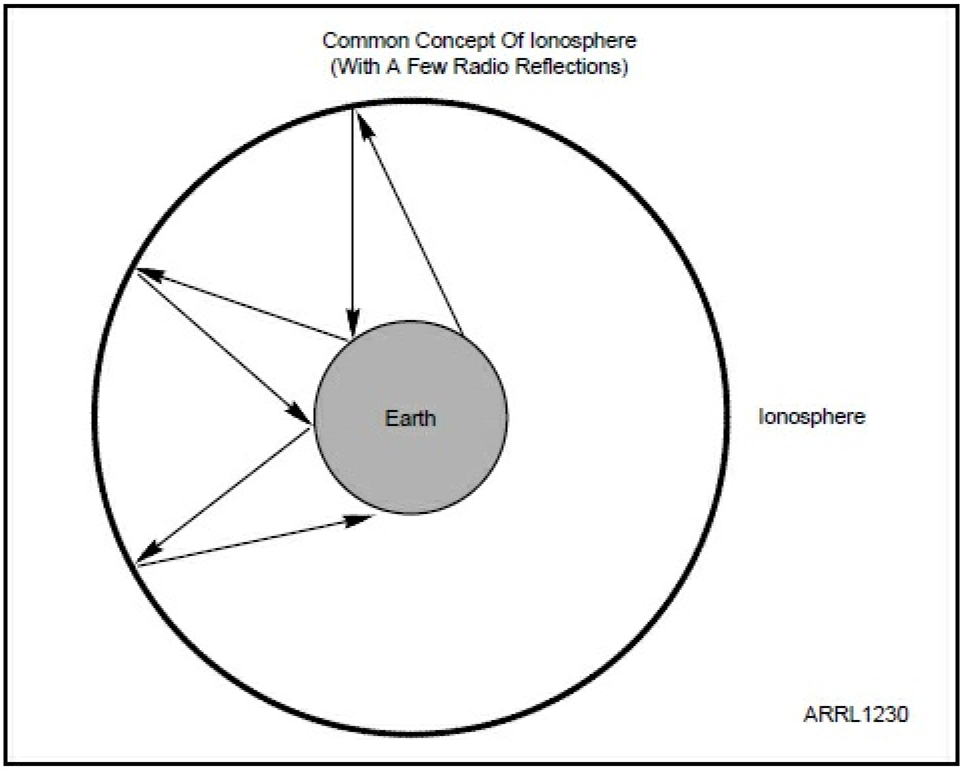


Figure 6.1 — A lot of misconceptions about radio propagation arise simply because we have the proportions wrong. A typical presentation of the ionosphere is shown in this figure.

To keep things manageable, let's first look at some things we *do* know about the ionosphere, in no particular order.

The ionosphere is *huge.* The unassisted breathable atmosphere around the Earth compared to the planet itself, is roughly proportional to the paper skin around the outside of an onion. It goes from ground level up to about 5 miles. In contrast, the ionosphere….that part of the atmosphere responsible for doing interesting things with radio, is about equal in volume to one third of the Earth's volume. It starts at about 60 miles in altitude and ends at about 500 miles. About 500 miles is the useful upper limit for HF/VHF as the ionosphere decreases above the F2 peak. This height can depend drastically on the latitude, as well; the lines of demarcation between and within layers is a bit fuzzy.

The ionosphere is incredibly sparse. Despite the tremendous volume of the ionosphere, there is almost nothing there. Physicists at the EISCAT facility in Tromso, Norway, not long ago calculated the mass of the entire atmosphere…it is on the order of *one* metric ton. Get that. One Metric Ton. You could compress the entire ionosphere and roll it into the corner of your garage with a hand truck. The density of the outer limits of the usable ionosphere is on the order of a *quadrillionth* of the atomic density at ground level. **Figure 6.1** shows the common concept of the ionosphere, and **Figure 6.2** shows a much more realistic model.

The ionosphere has weather. Since the ionosphere is *made* from "normal" atmosphere, it basically rides on top of it. It is not surprising that it has all kinds of weather-like effects — storms, turbulence, even tides. We should not be surprised that it's unpredictable at times. We should be more surprised that it's *ever* predictable!

On the average, the ionosphere has about the same number of ions as it does free electrons. This may seem fairly obvious, but it is important point to remember that if it were not the case, "normal" ionospheric propagation would *probably* not be possible.

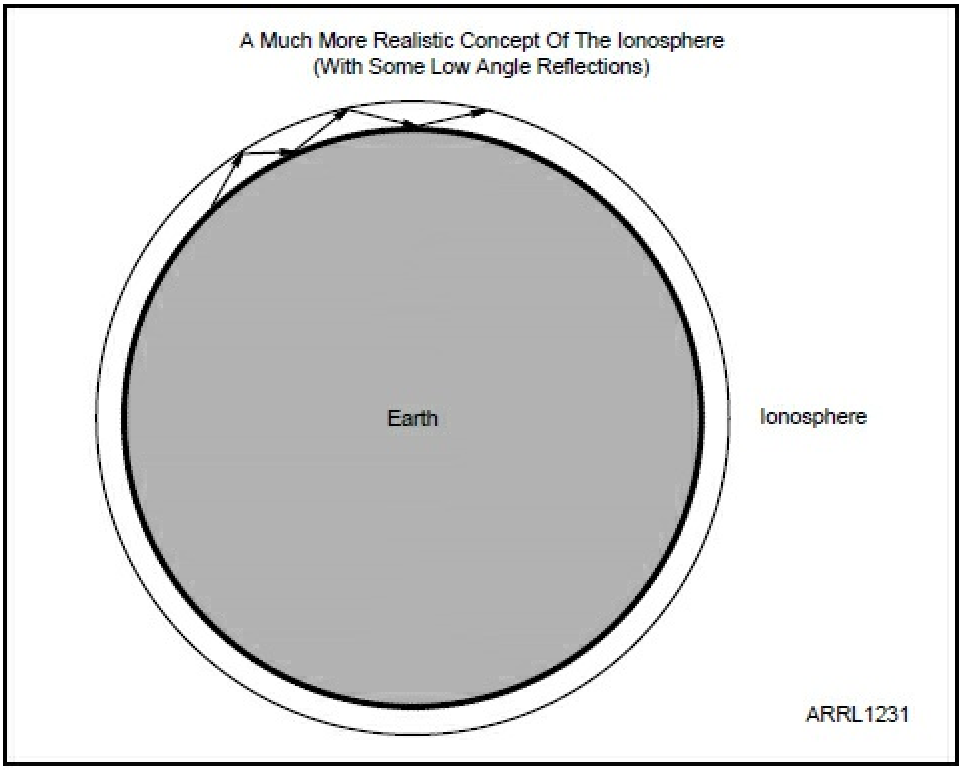


Figure 6.2 — This shows a much more realistic scale of the relative distances involved in long distance radio propagation. Although the ionosphere is a *sphere*, over small distances it appears much more like a horizontal ceiling. This factor actually simplifies some calculations, especially when multiple "hops" are involved.

The Earth's magnetic field has a *tremendous* effect on the ionosphere. This aspect has been greatly underestimated in most books on radio propagation, as well as propagation modeling programs. As we will see, it is the Earth's magnetic field that accounts for the vast majority of "oddball" ionospheric phenomena, and it's not as odd as we've often believed.

My ionosphere is not the same as *your* ionosphere. This, again, should be obvious to the active ham, but the fact that the ionosphere changes rapidly in time and space is often neglected in propagation literature.

How to Make an Ionosphere

Perhaps it's more important at this point to ask *why* make an ionosphere. Why is the ionosphere special? What makes it different from some of the other spheres, say the *troposphere,* we might encounter with radio?

Here is where you will appreciate why we started out with some principles of particles in the first chapter. The *neutral* atmosphere — the atmosphere that consists of oxygen, nitrogen, and a few other rarer gases — is totally invisible to radio. Well, not *totally¸* as we will learn later, but *nearly* so. This is because in a neutral atom (or molecule) there are equal charges of positive and negative. There is no *net* charge for an intercepting radio signal to interact with. The forces will always be equal and opposite. No matter how much radio power you have available, you cannot wobble a neutral particle around with it. A neutral's charges are pretty well self-absorbed within the neutral; they really don't care about external fields and forces. This truth is modified slightly when the wavelength of the intercepting radio wave approaches the physical dimensions of the neutral itself, but this doesn't happen often in Amateur Radio. Exotic technologies such as "laser induced fluorescence" can "drill down" into the inner workings of a neutral…but this is not "normal" radio behavior.

In order to prepare a particle to respond to a radio signal, we have to separate the electrons from the protons…or more generally, from the *nucleus.* This process is called *ionization.* This takes a tremendous amount of energy…at least in comparison with what we can supply from on Earth with our measly HF transmitters. The primary ionizing agent is radiation from the Sun, primarily in the form of ultraviolet (UV). X-rays can also cause ionization, but this tends to be actually disruptive, from a radio standpoint.

When an atom (say nitrogen) is ionized by a UV blast, an electron gets slapped out of orbit, and moves away from the remaining nucleus (which is now a positive *ion*). However, it doesn't move too far, because the nucleus is now *positively* charged, and has an electrostatic attraction for the off-slapped electron. This is actually a very good thing for radio. If by some magic, we could simply throw a bunch of free electrons up there without any remaining ions, the electrons would do what they do naturally: repel each other. They would scatter off into the far reaches of the atmosphere, not doing any of us any good whatsoever.

So, even though the free electrons are the end result we're looking for, without the ions to keep them in check, they wouldn't do us much good. That's why we call it the ionosphere, not the "electronosphere." Even though the ions and electrons are approximately equal in number, it's the *ions* that hold everything in a *sphere* rather than an amorphous cloud of unruly electrons.

Now, you're probably muttering to yourself, "Okay, but what keeps the *ions* in a sphere?" The answer would be the same as the answer to "What keeps the *atmosphere* in a sphere?" That would be gravity. Remember that a proton weighs 2000 times as much as an electron. And, the nature of gravity is that it's totally isotropic. It always pulls perfectly symmetrically toward its center, which is why the atmosphere is pretty close to a perfect sphere…minus local disturbances. So, you can see that the whole system seems to be custom made specifically to make radio happen. At least some of us hams like to think so!

Well, there are a lot of things that can go wrong in the process, and even when they all go *right*, it can be pretty complicated. But it can also be modeled pretty accurately.

Let's expand our model from just one or two ionized particles, and replace it with untold trillions. We'll marinate our atmosphere with a nice rich bath of UV rays, so that a large proportion of the atoms are ionized. If *all* the available atoms were ionized, we would call the "target" *fully ionized*, but that only happens rarely, and in relatively small regions. As we will discover, it is the *mixture* of ions and neutrals (and, of course, free electrons) that give the ionosphere its "personality" at any given time.

A few self-evident truths should emerge from this, one being that the greatest amount of ionization should occur on the side facing the source of UV (the Sun). And the highest ionization levels should occur at local high noon, all other things being equal. And, not surprisingly, we find that there is a high degree of correlation between the height of the sun and the maximum usable frequency (MUF), a parameter we will discuss thoroughly in a little while.

A good deal of the unpredictability or "charm" of ionospheric propagation is due to the *mixture* of neutrals and ions, and their interactions. We will find that some modes of propagation such as sporadic E are much more straightforward in some respects, where the reflection is largely due to highly concentrated clouds of electrons.

The mixture of ions and neutrals is largely (but not entirely) responsible for the *frequency* *dependent* character of the ionosphere. But before we go too deeply into this aspect, let's look at a few more self-evident aspects of the ionosphere.

We know that air pressure is dependent on altitude; the density of particles decreases as you move upward. But since air is *compressible*, the density decreases in an exponential manner with altitude, not linearly. (Conversely, if you're a scuba diver, water pressure is nearly directly proportional to depth). The net result is that air gets very thin very fast as you move above the troposphere, which is that part of the atmosphere we normally consider the "weather region." Since ions *come* from air, it's fairly obvious that the ion density would decrease with altitude in a similar manner. Well, this is true to some extent, but not entirely.

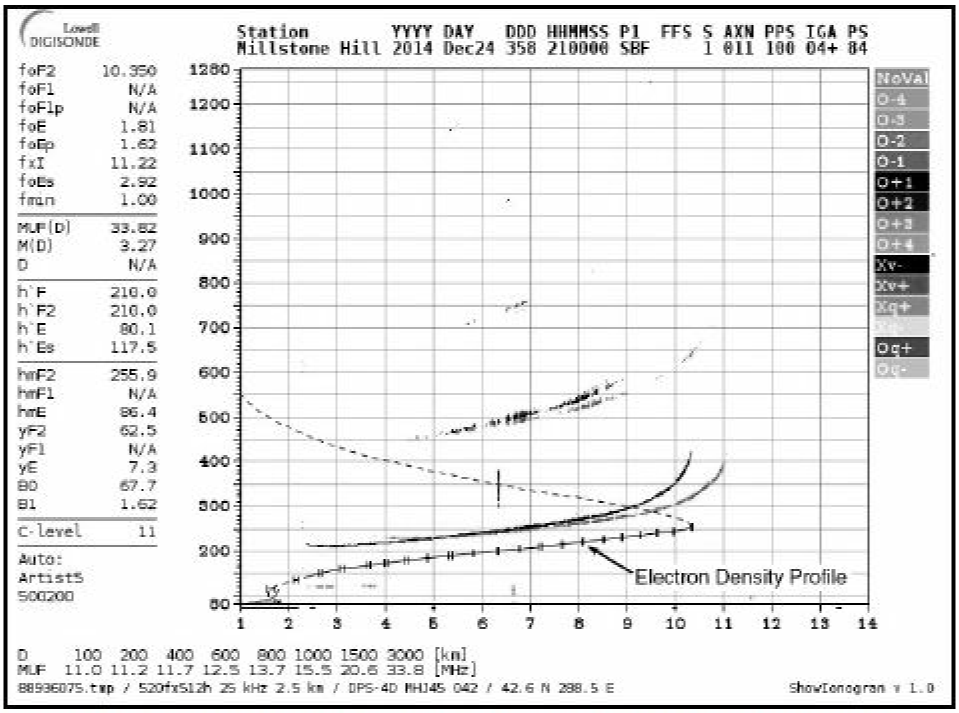


Figure 6.3 — The Electron Density Profile is a prominent feature of any ionogram. It is the somewhat haystack-shaped curve lying on its side, with the peak just above 250 kilometers, in this case. Above the peak the curve is dashed rather than solid, because anything above this height cannot be measured with ground based radar, but must be estimated mathematically. This region *can* be measured with "topside sounders" which are occasionally deployed by scientific rockets. This information is seldom readily available, however, so we must rely on ground-based sounders.

At high altitudes where the pressure is very low, it's very *easy* to create ions. It doesn't take very high levels of UV to do the trick. On the other hand, there isn't very much there to ionize in the first place. As we move downward, we find a lot more atoms and molecules available for ionization, but we find that these are more difficult to ionize because of the increased pressure. What we have, therefore, is a conflict between two parameters: ease of ionization, and volume of ionizable particles. The result of this balance gives us the all-important *electron density profile* seen in the ionogram in **Figure 6.3** (more on ionograms in a later section). We see that there's a definite peak in the concentration of free electrons at a specific altitude, typically around 250 kilometers at local high noon. This height can be highly variable, as well as the actual concentration at that height, which adds to the interest of ionospheric propagation.

Another fascinating aspect of this, one that has only recently been confirmed by *in situ* sounding rocket measurement, is that the *ion* density profile almost exactly tracks the electron density profile. This strongly suggests that the electrons never stray very far from the ions from which they are separated. Of course, we already described how something like this would be necessary in order to have a well-defined reflective layer, rather than willy-nilly globs of random electrons floating around.

We need to re-emphasize that at no point in the ionosphere do we have nothing but electrons, or nothing but ions. Every electron comes from somewhere, and that somewhere is a nearby atom or molecule. The electron density profile shows us the *concentration* of ions and electrons at a certain altitude, but these are still mixed in with neutrals.

A Truly Magnetic Personality

No explanation of the ionosphere is truly possible without considering the Earth's magnetic field. For decades, most amateur literature has ignored the importance of the Earth's magnetic field, and as a result most hams have attributed many weird behaviors of the ionosphere to just the nature of the beast. As it turns out, most of the bizarre things we see in the ionosphere are explained, or at least simplified, when you consider the Earth's magnetic field.

The primary effect of a magnetic field on a plasma (a region of ionized gas) is that the plasma becomes *birefringent*. This means there are two refractive indices, and they exist simultaneously. These two refractive indices are dependent on the *direction* of radio signals passing through the plasma. A radio signal passing through a birefringent medium becomes circularly polarized, as well, with one mode (the *ordinary* or O-mode) ray spinning one direction and the *extraordinary* or (X-mode) ray spinning the opposite direction. This has profound effects on radio propagation. Here are just a few of the more common ones:

• The ionosphere is no longer reciprocal. The old radio "wisdom" that there is no such thing as one way propagation is rendered invalid. In fact, as we shall see, in the strictest sense, one way propagation is the *norm.*

• Great Circle routes are increasingly meaningless as you approach the Earth's magnetic poles. This is more than obvious to anyone who has ever operated in Arctic regions!

• All ionospherically refracted signals become either X or O waves, most frequently *both*. There is no other option, unless the entire path happens to be precisely along the magnetic equator — a virtually impossible condition.

• X and O rays can take radically different paths, although this depends on frequency. At the higher end of the HF bands (say 28 MHz), the O-wave and X-wave paths are very similar (as is absorption). At the low end (say 1.8 MHz), they are not. Compare the ray traces from *Proplab Pro* propagation software in Figure 6.3 (10 meters) and Figure 6.4 (160 meters). There can be a tremendous advantage, therefore, to using circularly polarized receiving antennas.

Specifics About Species

In plasma physics, the primary science behind the ionosphere, one uses the term *species* to describe the actual type of neutral or atom with which one is dealing. Most Amateur Radio literature doesn't spend a lot of time talking about whether we're ionizing oxygen or nitrogen, or some other molecule or atom. There's a good reason for this. It doesn't really matter too much. Since the main ingredient of radio interaction is the free *electron*, we really don't care where the free electron comes from. We'll gladly take a free electron or two from any species willing to give it up!

However, certain phenomena we observe, such as the differentiation between the F1 and the F2 layers are the result of different ion species. The F1 layer, being lower, consists of generally heavier species such as oxygen and nitrogen, that don't ionize very easily, and thus is not seen as frequently. The F2 layer, on the other hand, consists of helium, hydrogen, and noble "neon light" gases which are much more easily ionized, and tend to hang around longer after ionization, thus the F2 layer is available for more hours than the F1 layer. The F2 layer is primarily caused by the ionization of oxygen atoms, with recombination due to molecular oxygen and molecular nitrogen.

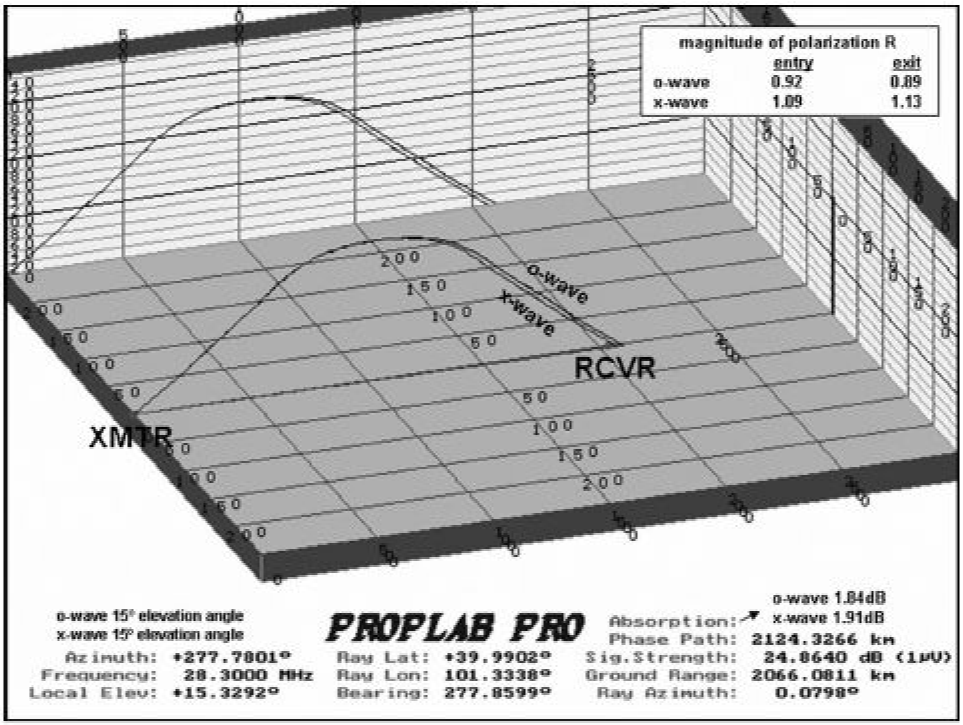


Figure 6.4 — As shown in this ray trace plot from Proplab Pro, at 28 MHz the O-wave and X-wave take very similar paths. [courtesy Carl Luetzelschwab, K9LA]

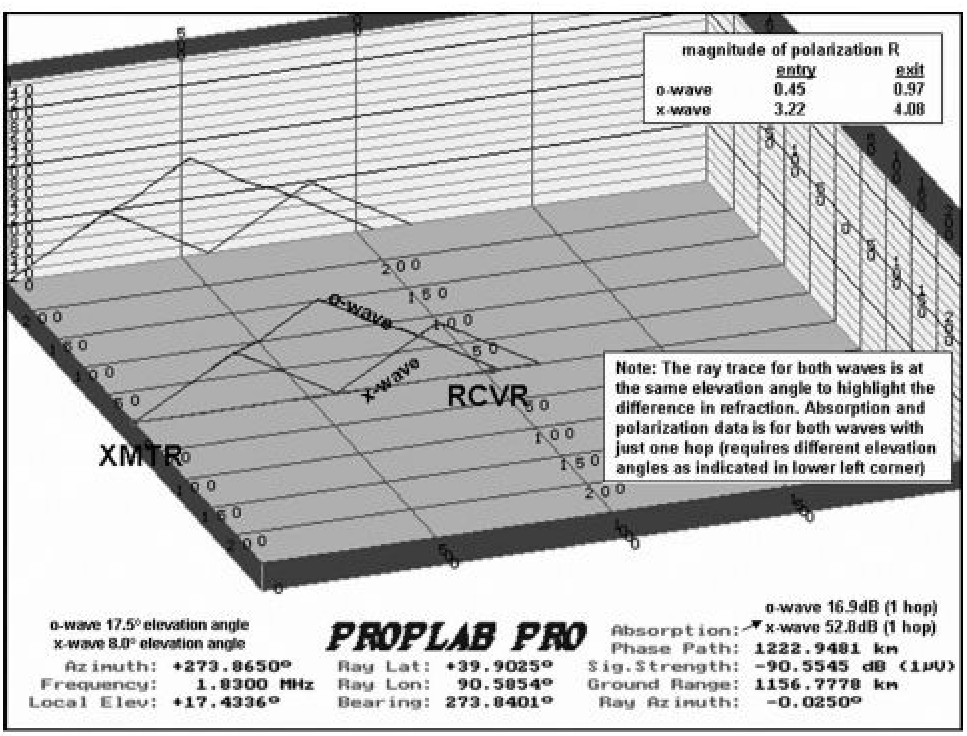


Figure 6.5 — Compare this *Proplab Pro* ray trace with Figure 6.4. At 1.8 MHz, the O-wave and X-wave take very different paths. Also note that in this case, the X-wave suffers significantly more absorption. [courtesy Carl Luetzelschwab, K9LA]

We also have *negative ions*, which, while generally short lived, have some effect on ionospheric behavior, but exactly what role they play is not completely understood.

Excite Me

In addition to the various *ionization levels* available to the different species, we also have different *excitation levels.* Excitation of a molecule or atom results when an electron is moved into a higher (larger diameter) orbit within the atom. Excitation requires a lot less energy than ionization, although both can occur simultaneously in the air. Visible light is emitted when an electron drops out of a higher level orbit into a lower one. Certain atoms such as nitrogen have *numerous* different excitation levels and modes, and these can account for a large variety of colors seen in an aurora, as the excited electrons return to their ground state.

One of the oddities of excitation levels is that they have a definite and well-defined *lifetime*, the time they stay in an excited state. This lifetime can range from milliseconds to minutes. Since an electron "dropping back" to its ground state is pretty much a random event, it may seem odd that there can be a defined excited lifetime. As it turns out, lifetime is a purely *statistical* phenomenon, and while the lifetime of an individual excited state is random, as a collective "gas" the lifetime turns out to be quite precise. The so-called "population inversion" that causes a laser to "lase" in synchronization is closely related to this excitation lifetime.

Now, as far as radio propagation is concerned, excitation levels have little noticeable effect. It is only when you start poking the ionosphere with optical frequencies that these excitation states become truly relevant. Interesting phenomena such as *laser induced fluorescence* are the result of various excitation states. Laser induced fluorescence has been used as an experimental communications mode, by the way, and is a wide open field for Amateur Radio experimentation, since we are effectively "licensed" for all frequencies above 300 GHz! Not long ago, the type of lasers necessary for performing this were beyond the reach of nearly every radio amateur, but with the incredible recent advances in laser diode technology, this is no longer the case!

Don't Skip This Part!

As we suggested earlier, radio amateurs have done some pretty amazing things with radio with some really rudimentary understanding of what's actually happening "up there." And we will never deny that a great deal of the romance of Amateur Radio is a result of the mystery and unpredictability of the ionosphere, as well as some of our sheer ignorance.

The phenomenon of radio "skip" has been observed for nearly a century, and many aspects of it are pretty intuitive…if we consider the ionosphere to be a mere reflecting "sphere." You don't need to know much (if anything) about ionospheric chemistry to have a lot of fun and effective Amateur Radio communications! But to make any real progress toward "advancing the radio art," we should learn all we can about this vast part of our world. For any radio amateur with a modicum of sheer curiosity, there's enough here to keep one occupied for a lifetime. Ionospheric propagation will probably never be fully predictable, but it can be largely *understandable.*

A Guided Tour

We can start putting a lot of pieces together by following a radio wave on its journey from a radio transmitter to its final destination. We'll describe the phenomena a radio signal is likely to encounter along its often circuitous path during fairly *normal* conditions, and then do the exercise again with a few more unusual factors included.

Let's say we want to transmit a radio signal between Palo Alto, California (the center of Silicon Valley) and Virginia Beach, Virginia (on the opposite coast, but near the same latitude) on 20 meters. There's nothing sacred about 20 meters, other than it's somewhere near the middle of the HF Amateur Radio spectrum, and has fairly decent daylight and nighttime characteristics. We'll use a moderate gain Yagi antenna to direct most of the radio signal in the direction of interest. We can use a Great Circle map to tell us, within relatively wide limits, where we should aim our transmitting antenna. Now, it may seem, at first blush, that if Virginia Beach is on the same latitude as Palo Alto, we'd want to aim due east. Well, this only works if you're right on the equator. In reality, the Great Circle path requires that you aim slightly *north* of due east if you want to hit a location at precisely the same latitude. (If you're below the equator, you will need to aim slightly south of where you think you need to go, in a similar situation). You can easily check this out with a globe and a length of string. But it's even better to use one of the many heading calculators available online (see Chapter 15, Software and Such). It's pretty simple geometry, at least in concept. The actual *calculation* of Great Circle paths is a bit more tedious.

Now, we need to keep in mind that a Yagi antenna is not a laser. On 20 meters, a typical three element Yagi may have a beamwidth of 60 degrees or so in azimuth (measured at the half power points), and generally an even greater beamwidth in elevation! This means your steering can be off by about 30 degrees before the other station even notices it.

But, for the sake of argument, let's assume you are steered right "down the bore" of the station in Virginia Beach. We still have a very "tall" beam in terms of elevation. Our "launch angle" can vary anywhere between horizontal and well over 60 degrees! We can certainly base some calculations on the *center* of the beam, center being the elevation of absolute maximum signal strength. And this is quite useful for determining the so-called "skip distance." If we know the height of reflection and the elevation of the center of the beam, we can, to a good approximation, determine where it's going to "land" after the first skip. We can use some straightforward geometrical optics to figure this out. And, we can even get a ballpark figure of how many skips it takes to get across the country *if* we make the assumption that the ionosphere is fairly uniform across its entire path. If it's high noon at the midpoint between Palo Alto and Virginia Beach, and the ionosphere is fairly calm, we can consider this a useful assumption. At least it will be *daylight* along the entire path.

Now, it should be fairly easy to visualize that if we look at the wide "vertical fan" pattern of a typical HF transmitted signal, after every "bounce" the fan gets wider! The range of launch angles after every return to Earth becomes greater. In fact, many shortwave broadcast stations take advantage of this fact, launching at fairly high angles right out of the chute as it were, so that there are no "dead" zones along the path between the transmitter and the target destination. We radio amateurs however, are more interested in getting a strong signal at a final destination, and not too concerned with picking up an "audience" along the way. For cross-country amateur operation, it's usually an advantage to launch at the lowest angle possible so as to require the fewest number of skips. Of course the ionosphere dictates what elevation angles are available over a path — and the antenna's job is to put the most energy at that angle, whether it be high or low. As we will explore later, far few hams pay any attention to the vertical launch angle of their signals, under the common assumption that they can't do much about it anyway. As we will see later, there's a lot more under your control in this regard than commonly thought!

Your Friend the Ionogram

If you operate HF, it's never too early to learn about the *ionogram*, one of the most useful tools in radio propagation and analysis. An ionogram is a graph generated by an instrument called an *ionosonde.* The ionosonde is a high frequency radar; it sends short radio frequency pulses straight upward (normally) and receives the returning signal after reflection from the ionosphere. Each pulse is transmitted at an increasing frequency throughout some or all of the HF frequency range. The X-axis scale at the bottom is the frequency, and the Y-axis scale is the reflection height. There is a wealth of information that can be derived from an ionogram, and we will look at some of these in detail as we move along.

The ionogram shown in **Figure 6.6** is from the Boulder, Colorado, Digisonde, conveniently located somewhat between Palo Alto and Virginia Beach. A number of Digisondes are active around the world, and you can most likely find one relatively close to your location. You can often find one near the midpoint of where you are and where you want your signal to go, which is where it will be most useful, but this ideal state isn't necessary. You can learn a lot by looking at any ionogram.

Well, let's take a look at this particular ionogram in Figure 6.6. As you can see on the X-axis, we have a scale running from 1 to 17; this is the range of frequencies, in MHz, that the ionosonde scans. Some Digisondes are programmed to cut off at 8 or 9 MHz. It typically takes a minute or so for a complete scan to be performed. Depending on the particular site, a new ionogram is generated every 15 minutes or so.

Along the Y-axis, we have a scale running from 60 to 700; this is the reflection height in kilometers. The Y-axis is really a time of flight, and the reflection height (let's call it the "virtual height") assumes the wave travels at the speed of light. This can also be reprogrammed to show lower reflections, but 60 km at the bottom is a very typical configuration.

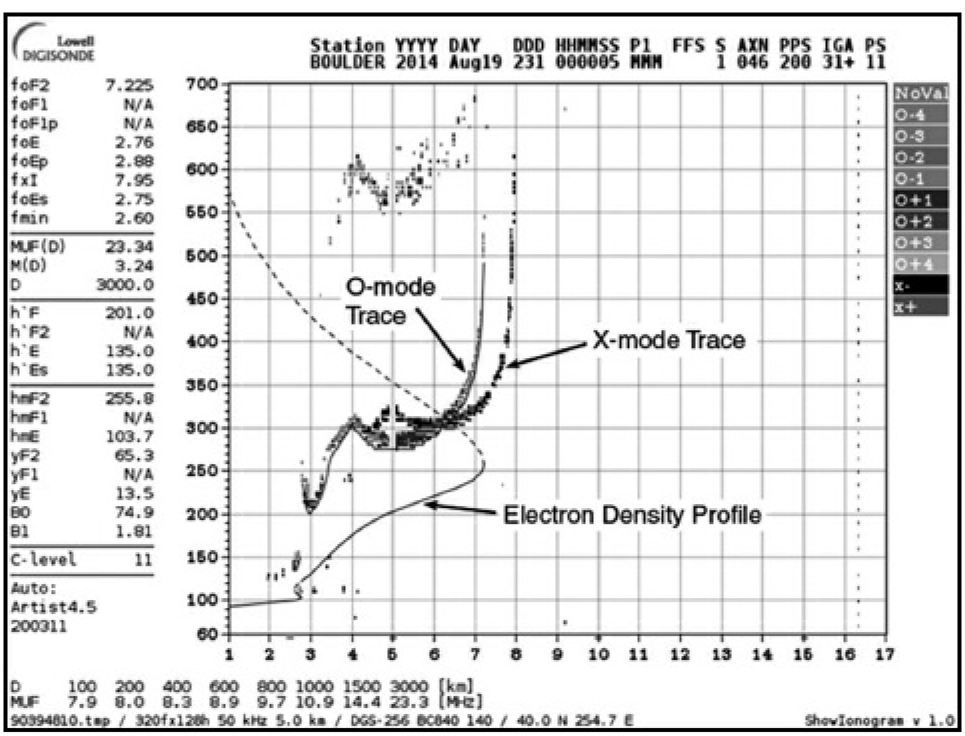


Figure 6.6 — A typical ionogram. See the text for a detailed discussion.

Now, let's look at the actual reflections, the plots we see on the graph. There are actually two traces of interest, the O-mode trace and the X-mode trace. Remember how we described the magnetized ionosphere as being *birefringent?* These two traces show this phenomenon clearly. The red and green traces are separated by means of two circularly polarized receiving antennas, one clockwise and one counterclockwise, while the transmitting antenna is *linearly* polarized.

Let's look at the O-mode trace first (the trace on the left here; it is in color on the actual ionogram). Look at the curve near 270 km in altitude, where the curve is fairly horizontal, between 4 and 6 MHz. (There appears to be a "cloud" up around 570 km. What's really happening is the 4 MHz energy is going up, coming back to Earth, going back up and then coming back down. Since we're really measuring time of flight, it shows up as twice the altitude since it's twice the time of flight.) We always want to look at just the first reflections, which give us the "real" information we're looking for. Now notice how the curve tends to increase with increasing frequency, and goes straight vertical at just over 7 MHz. This is called the *critical frequency*, and is the maximum frequency that will be reflected to Earth for a *vertical incident* signal. As we will learn, this is a worst case condition; the *maximum usable frequency* for lower angles is greater than the critical frequency, but related.

There is a column of "stats" to the left of the graph, of highly useful information, The first item is foF2, which is the critical frequency for the O-mode ray. In this case it's 7.225 MHz, which is what corresponds to the frequency at which the curve "goes vertical." This is the critical frequency of the F layer. There is also a critical E layer frequency (f0E) on the ionogram at 2.76 MHz. The E layer is easily recognized because it looks so flat compared to F layer refraction. The E layer critical frequency is a *low* frequency limit, not a high one.

Now, direct your attention to the rows below the graph labeled, "D and MUF." D is distance per hop in kilometers, and MUF is maximum usable frequency. Notice that the MUF goes up proportionally to the D.

It shouldn't be too surprising that the distance per hop is a function of the vertical launch angle and the reflection height. Some simple line-of-sight geometry should make this pretty obvious. Now, we can do a little more trigonometry and determine the actual launch angle based upon reflection height and hop distance. But here's the interesting part that shows the reflection process is actually refraction. The maximum usable frequency is dependent on the angle of the wave penetrating the ionosphere, and in this case it ranges between 7.9 MHz (which is essentially the critical frequency for vertical incident waves), and 23.3 MHz for very low angle radiation. For horizontal rays, the skip distance is about 3000 km, which means you can cross the country with just a couple of hops at 23 MHz…pretty good conditions for 15 meters.

Returning to the left column, we can drop down to FxI, which is the X-mode critical frequency. It's a little more than 500 kHz above that of the O-mode ray, at 7.95 MHz, seen clearly on the O-mode trace as well. A few lines down, we have MUF of 23.34, which again, is the case for a horizontal launch. A few more lines down, we have hmF2, which is the height of the reflection at the "normal" region of the O-mode. The flatness of this region can vary considerably, but it's a good first order approximation of the reflection height.

Virtual Reality

One of the unnecessarily confusing figures you'll find in this business is the difference between "virtual" and "true" height. Because the ionosphere is a refractor, not a reflector, the speed of radio is not constant through its journey. As the radio wave penetrates the ionosphere, it gradually slows down, so it appears that the reflection point is a lot higher than it is. (Since time of flight is the only indicator we have on Earth for the reflection height, we have to take this into account). The "real" or "true" reflection height; that is, the point where the wave velocity slows to zero, reverses direction, and accelerates down toward the surface, is a bit lower, depending on many factors. When calculating skip distances, it's the virtual height that really matters, because it locates the "reflection" point in the ionosphere. We will address some of the other parameters later, especially with regard to E and sporadic E propagation.

Electrons

Last but not least, we have the *electron density profile.* This is the black squiggly line that starts at about 95 km in altitude and runs up to around 575 km. What this line shows is the relative number of free electrons versus altitude. The horizontal scale is not shown, but it is generally measured in electrons per cubic centimeter or some such. Note that it's not the actual numbers that are important but the relative density. Notice there is a well-defined peak at 250 km. This is where you have the greatest number of free electrons. The peak of the F2 layer electron density is at the F2 layer critical frequency — from that it's easy to calculate the number of electrons. This is called the *critical height*, and it's also the height above which ground based sounding is impossible.

Now, see the dashed part of the line above the peak of the electron density profile. It's interesting to know that we can get no reflections above this altitude. The reason the upper part of the line is dashed, is that the electron density cannot be directly measured above this point, so it must be extrapolated. Fortunately, it is well known that the electron distribution follows a parabolic curve above that point.

Notice also the tiny little bump in density at around 100 km. This is E or sporadic E. (Note that the following text is about sporadic E, not the normal E layer.) These electrons are generally caused by either ablation (burning up) of micrometeors entering the atmosphere, or local lightning strikes. It is likely also that wind shear effects may cause these electrons to appear. What makes these electrons very different from "normal" ionospheric electrons is that in this region we do have almost nothing but electrons, without nearby associated ions. Because we don't have the stabilizing effect of ions in this area, the electron clouds are, almost by necessity, short lived, or "sporadic." Another feature of sporadic E (which will be shown in a different ionogram) is that there is no discernible frequency dependence…the reflection curve is remarkably flat. This speaks strongly of true reflection instead of refraction. We go abruptly from a region of no free electrons to nothing but free electrons when sporadic E is in play. There is still a great deal to be learned about sporadic E, and it is a wide open field for experimentation.

Layer by Layer

We've described how the presence of ions guides the ionosphere into fairly well-defined layers, which serve to make the whole region suitable for radio propagation. Let's talk a bit more about these layers and their unique characteristics. There's really something for everybody "up there" whether you operate "from dc to daylight" or never venture out of the confines of your 2 meter handheld.

Ionospheric layers are sorted by letters of the alphabet, beginning with the D layer. Perhaps you're wondering why there's no A, B, or C layers. Well, actually there are, but those are not actually ionospheric layers, but the troposphere, the stratosphere, and the mesosphere, respectively. Those particular spheres are not part of the ionosphere because they are not ionized, curiously enough. This is not strictly true, however. A lightning bolt creates very high levels of ionization, right down to the ground, but these are confined to very small regions and nothing resembling a worldwide *sphere.* And this is not to imply that there aren't some very interesting things happening near ground level. But for the most part, we begin our radio adventure at the D layer, which lies at an altitude between about 60 and 90 kilometers.

An Absorbing Subject

Most of what happens at the D layer is actually counterproductive to radio propagation. It is called the *absorption* layer, because when it is active, during daylight hours, lower frequency HF signals and medium wave signals are greatly attenuated. It is D layer absorption that is responsible for AM broadcasting to be strictly ground wave during the day. Such radio signals would be refracted from the upper ionosphere, except for the fact that they never get past the D layer in the first place. D layer absorption is actually strongly in effect up through the 80 meter band, with somewhat lesser effect on 40 meters. For this reason, the "low bands" are essentially nighttime bands, most effective when the D layer has dissipated.

The actual *process* of D layer absorption is rather complex and counterintuitive in comparison with the more "beneficial" layers. One could understandably conclude that if a little ionization is good for radio propagation, a lot of ionization is better. And indeed, it takes a lot more energy to cause ionization to occur at the low altitude of the D layer. Well, there can be too much of a good thing. D layer absorption is caused by frequent *collisions* between electrons, where they don't have the degree of freedom that they do at higher altitudes. There are a lot of complex *resonance* phenomena happening too, which contributes to the losses at low frequencies. Higher frequencies radio signals tend to slip by the resonances, unperturbed to a large degree.

D layer absorption phenomena can be particularly strong in polar regions where the Earth's magnetic field lines converge. *Cyclotron* resonances are prominent in these regions, as electrons oscillate about the magnetic field lines. The program *Proplab Pro 3* is one of the few ionospheric programs available that addresses cyclotron resonances. See Chapter 15, Software and Such.

E is for Excellent

Once you manage to make it past the D layer, radio becomes a lot of fun. The next level of interest is the E layer. For many hams, E is synonymous with sporadic E; in fact, you hardly ever hear about non-sporadic E in typical Amateur Radio discussion. The fact of the matter is that E is a region and *sporadic E* is a *phenomenon.* Sporadic E happens in the E region, but there is a lot of other stuff happening there too!

The E layer lies between about 90 and 120 kilometers in altitude. It's a rather thin region, though it might be hard to fathom 30 mile slice of sky as being thin! In terms of the entire ionosphere, it is a mere onion skin, however.

The E layer consists mainly of ionized molecular oxygen, (O_2), ionized primarily by soft (longer wave) ultraviolet. Most curiously, "normal" E is sort of a backward absorption layer; at low angles of incidence, it absorbs signals *above* about 10 MHz and refracts those below 10 MHz quite effectively. However, because the layer is so thin, it's quite easy to "miss" low angle E-layer refraction. In addition, because it is thin, one is unlikely to observe *dispersive* effects. Above the absorption frequency, the reflections from this layer are relatively independent of frequency.

While "normal" E propagation subsides quickly after sunset, *sporadic E,* on the other hand, can occur at any time of day or night. This is because the ionization process is not due to a steady bath of UV, but rather by very intense *localized* events. While the jury is still out on this, there are three most likely causes of this intense ionization: ablation of meteorites (micrometeors), lightning storms, and wind shear. All these phenomena can generate vast concentrations of free electrons. Wind shear is the most intuitive, because it is most analogous to the buildup of static electricity by the rubbing of dissimilar objects. Once you have blobs of electrons available, you have highly reflective mirrors. Sporadic E reflections are notable because of their highly *specular* or mirror-like reflections. Using HF radar, sporadic E reflections look like solid metallic objects, with well-defined boundaries. These blobs of electrons are effective reflectors of any radio signals whose wavelength is significantly less than their dimensions, up to the upper VHF range, where absorption becomes an issue again.

The F Region

Above the E layer we have the F region, which actually consists of two layers, ranging from about 200 km to 600 km in altitude. It is by far the "biggest" region of the ionosphere, by volume, and as such has the widest variation of propagation within its domain. It is in the F region where we observe wide radio frequency *dispersion*, the frequency dependent propagation of HF signals. Because the F region is so thick, HF radio signals spend a lot of time there in transit.

During the daytime, in seasons of high solar activity, the F region can separate into two distinct layers, the lower F1 layer and the higher F2 layer. The F2 layer remains considerably thicker, however, and for the most part, the F2 will play the predominant role in worldwide HF communications. One can spend a long "career" in Amateur Radio without actually noticing F1 activity. When it exists, however, the F2 layer is clearly visible on an ionogram, and can be responsible for very strong "short skip" signals.

Figure 6.7 shows an ionogram with foF1 (the F1 layer critical frequency) reported at 4.43 MHz. In this image, foF1 corresponds to an inflection point in the electron density profile — it's not another peak like the F2 layer (foF2 at 10.6 MHz).

As the Sun goes down, the F region "retreats" from the bottom up. The upper extreme pretty much stays fixed, while the region gets thinner. The F1 layer, naturally, goes away first, and the band "goes long" as the reflection height increases in the F2 layer. What's truly amazing is just how *fast* this retreat can be. Using Doppler ionosonde methods, the reflection height can be seen to move upward at more than 700 meters per second, right after local sundown! This can continue for a half hour or so until the reflection height settles in at its nighttime altitude.

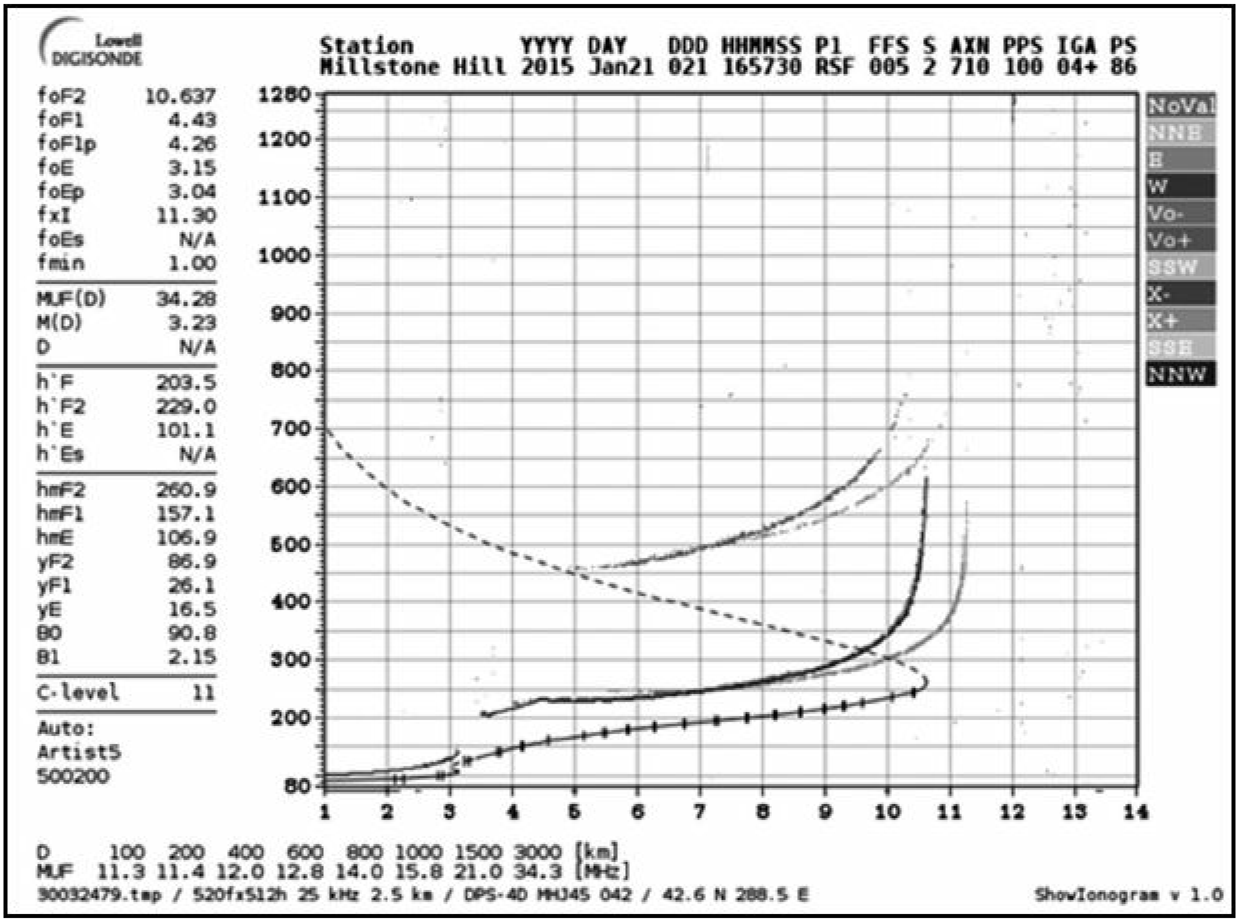


Figure 6.7 — This ionogram shows foF1 at 4.4 MHz (the small inflection in the electron density profile) as well as the obvious F2 layer peak at 10.6 MHz. [courtesy Carl Luetzelschwab, K9LA]

Now it shouldn’t take a rocket scientist to conclude that if the F layer retreats on the dark side of the Earth and advances on the daylight side, the entire ionosphere cannot be truly spherical! And, in addition, it has to have a pretty big tilt somewhere along the line. The net result of this is that if you launch an HF signal from the daylight side of the Earth across into the nighttime side, (or vice versa), the vertical angle of arrival will be much different than the vertical launch angle! Again, some simple geometry prevails.

This factor is not often noticed by hams, since, as we indicated earlier, very few are aware of or equipped to systematically deal with different vertical angles of arrival. As we delve into this more, however, we will find that, for the average HF operation, the control of antenna elevation may be every bit as important, or even more so, than azimuth control!

This chapter has been a fairly general overview of *expected* HF behavior, and for the most part will be confirmed by most experienced radio amateurs. In the next chapter, we will look at some of the more exotic behavior of the ionosphere, and methods of exploring and exploiting the same.

The Anomalous Ionosphere

In Chapter 6 we gave a rather broad overview of known ionospheric principles, employing some of the straightforward geometric optics discussed early in the book. Familiar tools such as Great Circle charts, solar indices, and countless thousands of person-years of experience have made it fairly simple to work most of the world at one time or another, with a reasonable level of reliability.

There are at least two reasons to explore the more bizarre aspects of radio propagation: to perform effective radio communications when conditions aren't "right," and for sheer scientific intrigue. Radio amateurs are a curious lot (or at least should be)! There is still a great deal we don't know about the ionosphere (and other spheres as well), and we are in a unique position to explore these realms. New modes of propagation have been discovered more or less regularly throughout the history of Amateur Radio, which in due time we have learned to exploit in practical fashion. For example, amateurs discovered transequatorial propagation (TE) and documented it in December 1959 *QST*. However, there is a lot of "stuff" up there which may be unlikely to serve *any* practical purpose for the foreseeable future, and yet should be of interest to any radio amateur with a modicum of curiosity. Anomalous propagation can range anywhere from the readily explicable *chordal hop* propagation discussed later in this chapter to the totally mysterious.

Low in the Tropo

There is probably no propagation phenomenon observed by more radio amateurs than tropospheric ducting, where VHF signals are propagated over far greater distances than they are "supposed to go." This is most likely due to the fact that in relatively recent history more hams start out using 2 meters than any other band. Nearly every active radio amateur has experienced tropospheric ducting at one time or another, often unwittingly. Not every unusual long distance VHF communication is due to tropo ducting, but this phenomenon is quite common…which may lead one to wonder why it's even called a *phenomenon.*

The troposphere is the part of the atmosphere we live in, and more or less extends from ground level to about 25,000 feet or so. Weather occurs in the troposphere, and as we know most weather phenomena are a result of atmospheric pressure changes. One of the results of high pressure upper tropospheric conditions is the potential of an *inversion layer* forming. This is where the temperature goes *up* with altitude over a small region. This gives the air in the region a well-defined refractive index gradient, which is ideal for refracting VHF and UHF radio waves.

Ordinarily, the troposphere is transparent to radio waves, as there are no ions with which to interact. However, even non-ionized species can have a dielectric constant different from vacuum, though normally very slightly so. This is more evident at VHF and UHF frequencies where the wavelengths are short enough to "wobble" atoms by a number of resonance effects. A neutral is normally unaffected by radio because the charges are so close together as to experience equal and opposite forces from an impinging radio wave. But remember, electrons are in *orbit* around their nuclei, so there *can be* a time differential between the force on the positive charges and the force on their negative charges (electrons). Under the right conditions, this time differential can have significant effects, such as the heating of water by microwave radiation.

True resonance effects are not likely to occur at VHF and UHF frequencies, but the waves are short enough to have slightly different effects on the electrons and the nuclei, thus causing a sort of "store and forward" action to occur. This results in radio propagation being slower than in a vacuum.

When a sudden density gradient occurs, this delay time is quite different over a relatively small vertical distance, which, similar to ionospheric refraction, gives us a prismatic effect for radio. Under extreme tropo ducting events, radio signals can propagate well over 1000 miles. It would be wrong to call this a "skip distance" however, because, obviously if the "reflection" point were no more than 25,000 feet (the upper height of the troposphere), the skip distance should be only a handful of miles, based on normal "geometric" calculations. However, this refraction occurs *continuously* along very great distances and at *very low heights*, from a radio standpoint. Tropo ducted signals take low, grazing paths along the bottom side of the inversion layer. It truly is amazing how well this process can work, with nearly no losses occurring under some conditions.

Like many other phenomenal events, one can miss tropo ducting events if you aren't looking for them. One of the best tools we have for predicting when "tropo is in" is VHF broadcast television or the FM broadcast band. (This is one reason not to scrap your "free air" TV set in favor of a satellite dish or cable.) You might try monitoring channel 2 or 3 (whichever is not in use in your area), as these are pretty sensitive frequency ranges for tropo. If you see far away stations appear in living color on your TV set, it's time to fire up the 6 meter rig! It should be noted that 6 meters is not only subject to phenomenal *ionospheric* propagation at times, but also tropospheric propagation…often simultaneously! This is not the case with 2 meters, which is well above any ionospheric propagation, frequency wise.

Part of the reason 6 meters is known as the "magic band" is because of its response to both ionospheric and tropospheric propagation, the only band that has this property (except for possibly a few countries that have 4 meters!).

Low Band Mysteries Abound

At the other end of the spectrum, but in the same "sphere" we have some unusual happenings, too. Theoretically, ground wave communications, such as nearby 160 meter contacts (or even AM broadcasting reception) should be utterly unaffected by tropospheric conditions of any kind, other than the moisture content of the ground caused by precipitation. For the most part, AM broadcasting within the ground wave region is astoundingly predictable. And yet, there is a volume of well-documented cases of "broadcast blackouts" where AM signals have entirely disappeared well within line of sight.

Several theories have been put forth to explain these outages, ranging from heavy particles such as *magnetic monopoles* or other exotic high energy oddballs passing through the Earth causing hugely absorptive ionization at the surface, to total phase cancellation from near vertical incident skywave (NVIS) propagation.

Likewise, truly bizarre propagation events on 160 meters (both good and bad) are observed too frequently to ignore, yet defy all "normal" explanations. Like many other intermittent phenomenon, we probably aren't going to learn much by merely speculating on the matter whenever these events happen to show up. The answers will come when we deliberately set out to *look* for these; a process which requires time, dedication, and well thought out instrumentation and methods.

Just as broadcast TV stations are valuable instruments for prediction of tropo ducting, AM broadcasting is a good reliable benchmark for exploring unusual low-band openings as well as outages.

One matter that is seldom discussed in Amateur Radio literature is *cyclotron resonance*, or *electron gyro frequency*, that is the natural resonant frequency of gyration of electrons (as well as ions) around the Earth's magnetic field lines. One of these resonances falls in the middle of the AM broadcast band, but there are weaker resonances in the 160 meter band, as well. What the exact effects of these gyro frequencies on Amateur Radio communication are as yet unknown, but they could be rather absorptive. Again, this is something that amateurs can readily explore with little more than some scientific discipline.

Distraction by Diffraction

Most VHF denizens are familiar with the phenomenon known as knife-edge diffraction; its occurrence is generally predictable enough to be a "non-phenomenon." Knife edge diffraction is a well-known optical principle, where when a wave encounters a sharp demarcation, such as a mountaintop (or a genuine knife edge), it bends abruptly at the boundary. This allows radio reception behind an otherwise opaque mountain ridge, for instance. For the most part, mountain ridges stay where they are, so if knife-edge diffraction is observed at any time in a particular location, it should be available just about any time.

It is not uncommon, however, for hams to talk about "seasonal" knife edge diffraction, which doesn't really seem to make much sense. A couple of possibilities arise. The most straightforward explanation is that the actual conductivity of the knife edge is variable…again, dependent on precipitation, or even nearby foliage. Another likely suspect is that the "complainant" is using a combination of knife edge diffraction and something else — such as troposcatter. When one of them goes away, such as troposcatter is wont to do, so does the signal.

No propagation phenomenon is mutually exclusive with any other. In fact, *most* practical long-distance radio propagation probably results from bits and pieces of many known phenomena.

The Scatter Matter

This brings us to the matter of tropo*scatter*, which, while related to tropo ducting, is not exactly the same thing. Troposcatter can occur from *any* kind of discontinuity in the troposphere, not just inversion layers, as in tropo ducting. However, in classical troposcatter, these discontinuities are microscopic…most frequently water vapor…which randomly scatter VHF signals in all directions, some of which, hopefully, are in the direction you want to go. While being far more lossy than tropo ducting, troposcatter is a lot more reliable. It always works, but it always works poorly!

While great distances can be achieved with legal power and high gain antennas (preferably at both ends), probably very few hams experience true troposcatter. In addition, while lower VHF will work for troposcatter, in general it works best above 2 GHz — high enough to be out of the range of the highly absorptive molecular resonance of water. In the middle of the last century, long-haul commercial and military troposcatter circuits were used in Alaska and elsewhere, where reliability took a high priority over efficiency. However, with state-of-the art weak signal digital modes, such as the "JT" modes, modern day troposcatter may be ripe for amateur experimentation. (More on the digital modes developed by Joe Taylor, K1JT is available at: **http://physics.princeton.edu/pulsar/k1jt/**.)

Underground Radio

As we established earlier, all ground-surface waves are vertically polarized…out of necessity. But what about *underground* radio waves — those that are launched entirely under the surface? Can such a thing even exist?

This depends on what your definition of ground is. If by ground you mean a solid volume of highly conductive medium — what we like to *think* the Earth is made of — the answer is no. Conductive volumes *short out* the E field of radio waves, and prevent the propagation through them. At *extremely low frequencies* where the *skin depth* is very deep, radio waves can propagate somewhat, but with extremely high loss. By the same token, undersea radio propagation can take place at ELF. But both of these modes are so dreadfully inefficient and lossy as to be utterly useless for Amateur Radio.

On the other hand, if by ground you mean "beneath the visible surface of the Earth," underground radio can certainly exist. For instance, here in interior Alaska, the ground conductivity is *so* poor that it basically appears as an insulator. Many hams in the bush operate just fine with an 80 meter dipole or horizontal loop *lying on the ground.* Likewise, in similar regions, underground antennas act pretty much as if they were embedded in a chunk of glass. Back during the Cold War, a number of underground ham antennas were marketed as being blast-proof, indestructible, and so on. In some places, they actually sort of worked, but for the most part were miserable failures. Some antennas such as *magnetic loops* are supposedly better able to work underground better, because the electric field component is less in proportion to the magnetic component, and thus less subject to the "shorting out" phenomenon. This is dubious at best, but still probably worth some experimentation.

Nonlinear Radio

Now we get to the most interesting and unexplored "region of the weird": nonlinear HF propagation. For the most part, ionospheric propagation is a *linear* phenomenon, in that the acceleration of charges is in direct proportion to the E field of the radio wave. Of course there are countless other factors at play, but for any given state of environmental conditions, the signal coming *out of* the ionosphere is in direct proportion to the signal going *into* the ionosphere. This is true even if we include the magnetic effects (X and O mode waves). We radio amateurs rely heavily on this linear property of the ionosphere. Things would be *really* strange if the ionosphere were something other than linear *almost all the time.*

Back in the 1930s, a phenomenon known as the Luxembourg Effect gave the first hint that the ionosphere was something other than a passive reflector. The Luxembourg Effect was the observation that another distant shortwave station was able to *cross-modulate* with Radio Luxembourg's powerful shortwave signal. The phenomenon was later observed on several other pairs of shortwave stations. Since Radio Luxembourg was thousands of miles away from the secondary station, there was no possibility of front end mixing or another similar circuit phenomenon. The mixing was clearly taking place in the ionosphere, but only at very high power levels. At the time Radio Luxembourg was the most powerful HF station in existence.

This phenomenon instigated a decades-long exploration of ionospheric physics, especially in connection with high powered radio, culminating in the construction of experimental facilities such as HAARP, Arecibo, EISCAT, and a number of other large "heating" facilities scattered around the world. The research still goes on.

The ability of the ionosphere to "mix" RF signals is a result of nonlinearities, just as in a common diode. There are several sources of these nonlinearities, but they all essentially stem from the fact that both electrons and ions are accelerated by passing electromagnetic waves…but at different rates of acceleration. It is the combination of these different accelerations that gives us the nonlinearity of the ionosphere that results in so many interesting phenomena.

The reason we don't normally see nonlinear effects is that the *lightest* ion is the proton, which is nearly 2000 times the mass of the electron, but with the same charge. This means we need at least 2000 times the E field to see the same acceleration from a proton as we do from an electron. Keep in mind that most ionospheric ions are *much* heavier than single protons!

The bottom line is that at legal power levels, we aren't very likely to be able to wobble an ion enough to see the results back on Earth. But it doesn't mean it isn't happening. It's all a matter of degree. As experiments at HAARP (High Frequency Active Auroral Research Program) and elsewhere have shown, there *is* no power threshold above which nonlinear action happens, and below which it doesn't happen. It's just that the results (that is, *converted* radio frequency re-transmissions) are usually far below any noise floor we can observe. With enough receiver sensitivity, and perhaps some digital manipulation, it is not out of the realm of possibility that any ham can be a HAARP!

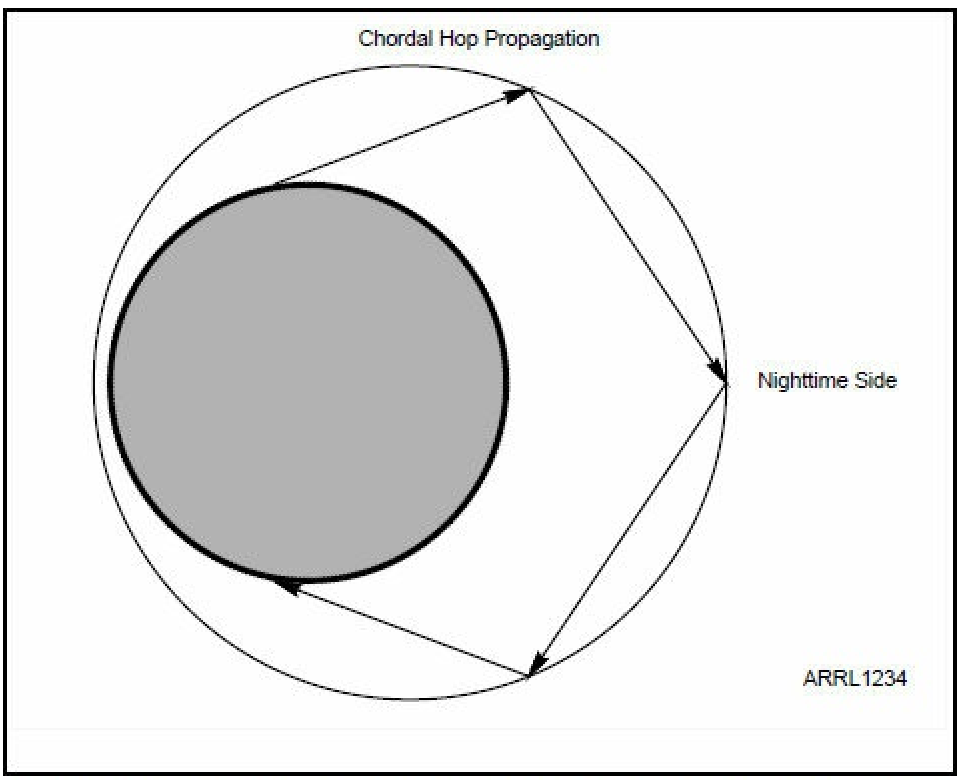


Figure 7.1 — The exception proving the rule. Certain conditions can cause the ionosphere to be much *higher* than normal, approaching the "misconceived" model shown in Figure 6.1. Strong solar winds can blow the ionosphere far from its "normal" course over the dark side of the Earth. This sets up a situation conducive to *chordal hop* propagation, where waves can bounce off the ionosphere multiple times with no ground reflection in between hops. Such propagation can be extremely efficient....but unpredictable.

Parametric Paradoxes

Anyone who's been on 80 meters for very long has, from time to time, experienced so-called "chordal hop" propagation (**Figure 7.1**). This is rather easy to explain. The ionosphere is somewhat elliptical, with the greater height being on the dark side of the Earth away from the Sun. There are two reasons for this. As described in the previous chapter, the dark side ionosphere retreats "from the bottom up" creating a higher ionosphere at midnight. In addition, the *solar wind* passing around the Earth from the daylight side drags some of the ionosphere along with it, tending to "pull" the ionosphere toward outer space. Both of these serve to form a rather lopsided ionosphere, where the path on the dark side is very high. With such an arrangement, a low-band radio wave can bounce off the ionosphere several times with no intervening bounces from the Earth's surface. It travels in *chords* around the ellipse.

Such signals tend to be unusually strong, following the path around the dark side much like tropospheric ducting. However, there are many reports of such signals being far stronger than can be accounted for by even free space conditions; in other words, it seems like they are actually being *amplified* along the way. Is this possible?

Answer: Possible? Yes. Likely? No.

This phenomenon, known as *parametric amplification* can be created consistently in plasma chambers, under tightly controlled conditions of magnetic field, pressure, gas species, frequency, polarization, and a few other minor details. Physicists can create parametric amplification on demand.

In the ionosphere, all the ingredients necessary for parametric amplification are there…sometimes. The real question is whether all the ingredients are there at the same time in the right proportion for "in the wild" parametric amplification to occur. This is *extremely unlikely.* However, as with countless other Amateur Radio discoveries, there's a difference between extremely unlikely and impossible.

Of course, it should be obvious that before one attempts to find the *reason* for a phenomenon, one needs to show convincingly that the phenomenon actually exists. Does it only *seem* like these signals are being amplified, or have any actual measurements been performed to confirm that the signals are stronger than equivalent signals in free space? If we can prove that the latter is the case, then we can collectively proceed to some *really* interesting experiments!

LDEs or Little Green Men?

Another equally puzzling phenomenon is that of *long delayed echoes (LDEs).* Far too many reliable observations of LDEs have been made for so many decades that it is impossible to just write these off. Something real is happening here. There are a few plausible explanations that don't require any extraterrestrial intelligence. Like parametric amplification, long delayed echoes can be created in plasma chambers using the known behaviors of *ion acoustic waves.* Again, the question is, are all the ingredients in place at the right time to create long delayed echoes "in the wild"? Again, another long shot, but not entirely impossible.

The answers will come when we set out to deliberately explore these events, not merely speculate on their sources when or if we happen to encounter them. Of course, this is easier said than done, but we have a lot of hams spread out over a very big Earth. Our odds may be better than we think!

Magnetic Personality

It is more than a little mystery as to why the role of the Earth's magnetic field on radio propagation has been all but ignored in Amateur Radio literature. The characteristic X and O mode waves have been thoroughly understood in the scientific community since at least 1940. In fact, in that year, *QST* did publish an article on the matter. The late Bob Brown, NM7M, began writing about X and O waves in his January 1992 *WorldRadio* column, and in subsequent issues of *WorldRadio* and *Communications Quarterly* magazine. Carl Leutzelschwb, K9LA, began writing about this in his September 1998 *Worldradio* column, and in subsequent publications and articles on his website (**http://k9la.us**). The topic of X and O waves didn't return to *QST* until my December 2010 article, "Gimme an X, Gimme an O….What's that Spell? Radio." It has been gratifying to see a number of authors pick up the baton in the past couple of years and give the topic the attention it so richly deserves.

We should mention that the existence of X and O mode propagation does not in any way negate what knowledge we do have about ionospheric radio propagation; rather, it just completes the story. See the sidebar "Taking Advantage of X and O Modes."

The Earth's Magnetic Field

The permanent magnet embedded in the Earth is much like any permanent magnet, only a lot bigger (**Figure 8.1**). Relative to the more obvious geographic poles, the magnetic poles are a bit out of kilter. Both the North and South magnetic poles are considerable distances from the "true" poles, and in addition, they are not antipodal (geographically opposed) to one another. There is a somewhat stronger overall magnetic field over the Pacific region, as both poles are located in that "quadrant."

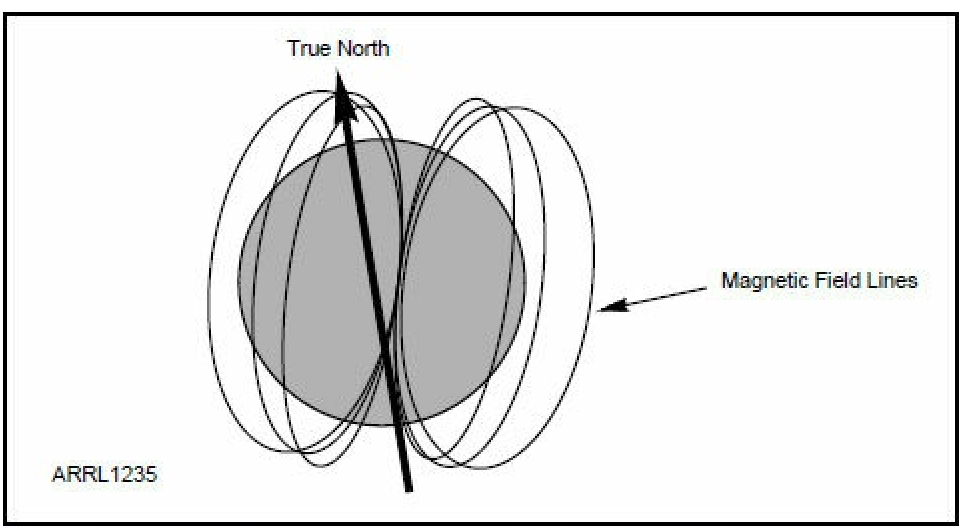


Figure 8.1 — The Earth's magnetic poles are displaced considerably from the true poles. In addition, the north and south magnetic poles are far from *antipodal*. This gives the Earth a rather lopsided magnetic environment. Note also that near the magnetic poles, the *field lines* are nearly vertical (normal to the Earth's surface at the entry points. Radio propagation is very different near the poles. Near the magnetic equator, the field is weaker, and more horizontally oriented.

The poles are located well beneath the surface of the Earth, most likely in the mantle. They tend to wander about a bit, and there is some indication that they are in the early stages of a pole reversal; this has probably happened a couple of times before in human history. (It's safe to say there probably weren't a lot of hams around the last time the poles "flipped" however). The poles are best defined as the place where the magnetic field lines are perfectly vertical. The amount of vertical angle of the magnetic field at any location on Earth is known as the *inclination.* (In Fairbanks, Alaska, this inclination is about 15 degrees). The more familiar *declination* is the difference in *azimuth* between True North and magnetic North. This wanders around a bit, too. If you still know what a magnetic compass is, you have probably adjusted it for the declination at your home location.

The south magnetic pole is currently off Antarctica between Adelie Land and Wilkes Land. Needless to say there are a lot fewer hams in the vicinity of the South magnetic pole than the North magnetic pole (where the highest concentration of hams per capita resides).

Like many such exercises, determining the actual depth of the magnetic poles is extremely difficult. You can't know the location on the surface without knowing the depth, and you can't know the depth without some very interesting triangulation from the surface! This is the sort of ambiguity that can drive a scientist nuts and the process of locating the actual position of these is far from complete… even if they didn't wander around! But we know this information with enough accuracy to be truly useful, at least for Amateur Radio purposes.

Magnets in a Vacuum

It's fascinating to realize that a stationary magnetic field in a vacuum — no matter *how* strong — has absolutely *no* effect whatsoever on an electromagnetic wave. Now, we should interject here that the Earth's magnet, while huge in volume, is actually very weak. The field concentration of a refrigerator magnet is many orders of magnitude more intense than the Earth's field, even right at the aforementioned geomagnetic poles.

Because of the relative feebleness of the natural terrestrial magnetic field, it's easy to assume that its effect on the ionosphere would be negligible. But, remember how we described how little ionospheric matter actually exists, that the entire mass of the ionosphere is on the order of *one metric ton*? The ionosphere is extremely thin. It doesn't take much to scoot it around a great deal.

Now, as we will see, it's the *interplay* of the magnetic field and the ionospheric plasma that gives it its particular influence over radio waves.

Being Biased

In the absence of a magnetic field, a plasma (a body or region of ionized gas) is extremely bilateral when it comes to radio waves. Or, put another way, radio propagation through a non-magnetized plasma is reciprocal. This is not to say that a uniform plasma has no effect on radio waves by itself, but its effect on them is entirely symmetrical or *isotropic.* It doesn't matter which way a radio wave goes through a uniform plasma; it always has the same effect, most frequently a slowing down of the radio wave.

In addition, a uniform plasma is essentially non-dispersive. It will refract radio waves, but it will refract them the same amount, regardless of frequency.

In the ionosphere, there are several factors influencing the uniformity of the plasma. First we have the pressure differential…air gets thinner with altitude. So we have a gravity gradient to the plasma. This is, naturally, pretty independent of azimuth; it is strictly dependent on height. We also talked about the electron density profile: the *ionization level* is dependent on altitude as well. The overall *vertical gradient* is, as we have shown, a complex function of altitude and ionization, but still extremely uniform with respect to *azimuth*.

If we apply a magnetic field to the plasma, a number of interesting asymmetries occur. First, a magnetic field tends to further change the plasma density, above and beyond the gravity gradient. Now, this may or may not be in the same direction as the gravity gradient. If you are directly over one of the poles where the magnetic field is vertical (parallel with the gravity gradient), what you will "see" will be just an exaggeration of the gravity gradient. Now, if we were to transmit a radio signal vertically, precisely in parallel with the field lines, we won't see much net result from the magnetic field other than an increased refractive index. In other words, the radio wave will travel slightly more slowly than in the case of the non-magnetized ionosphere.

Things get more interesting very quickly, however, once we transmit a signal that's anything but perfectly parallel to the magnetic field. In this situation, we have a component of the wave that *is* affected by the magnetic field, because it now cuts *across* it. The degree to which the magnetic field lines affect this propagation is directly related to how perpendicular the wave motion is to the magnetic field line.

But something else is happening at the same time. Once a linearly polarized radio wave begins to enter the magnetized ionosphere, it begins to split into two counter-rotating *circularly polarized* rays. These two rays diverge widely in both vertical angle *and* azimuth. One of these rays, the ordinary mode (O-mode) ray, travels along the field lines with not too much "weirdness." Its propagation path is much like that of a wave in a non-magnetized ionosphere…except for the fact that it's circularly polarized.

The other component or "characteristic wave" is the X or *extraordinary* ray. (Nobody said physicists could spell!) This ray follows basically *perpendicular* to the field lines, and has some more peculiar properties. The X-mode ray tends to be lossier, and it travels more slowly. It can also get "trapped" between the F1 and F2 layers, as if inside a waveguide, before eventually emerging at some location *far* from its associated O-mode ray.

As noted in Chapter 6, the divergence between X and O mode rays depends on frequency. In the 10 meter ray traces in Figure 6.4, the difference in absorption between the O-wave and X-wave is likely imperceptible in the real world. On 160 meters (Figure 6.5), the X-wave is considered out of the picture. Carl Leutzelschwab, K9LA, observes that from *Proplab Pro* ray traces, on frequencies below the electron gyro frequency (for example, on our experimental 630 meter band) the X-wave suffers less absorption than the O-wave. More investigation is needed here.

For a slightly different perspective on this process, refer to "Ionosphere Properties and Behaviors, Part 2" by Marcel H. DeCanck, ON5AU, in the July 2006 issue of *antenneX* (**www.antennex.com**). DeCanck correctly points out that nearly all propagation predictions are based on the O-mode wave behavior, which means we miss out on a great deal of interesting X-mode propagation.

So What

One could write a whole book about just this X and O mode business (and in the physics circles there *are* entire books on the matter). But hams are notoriously practical about such matters, and are correct in asking the question, why do we care about X and O modes? Does knowing of their existence have any effect on how we operate? Yes indeed! Let's look at the fundamental implications of X and O mode waves.

Not so Great Circle Paths

One of the most direct effects of the magnetic field is that Great Circle paths are increasingly more meaningless (or decreasingly meaningful) as you approach one of the magnetic poles. In the northern hemisphere, the O-mode ray will diverge toward the North magnetic pole, while the X-mode ray will diverge toward the magnetic equator. Anyone who has ever operated in interior Alaska will testify as to how frustrating it is to figure out where to steer your beam antenna, not only for over-the-pole signals, but also from signals arriving from the "Lower 48." Now, while this divergence is most extreme in polar regions, it exists to some degree everywhere. Unless the entire path is along the magnetic equator, there will always be some X-O divergence.

But What Can We Do About It?

This is a great question, with a profound answer. If you're a typical DXer, you are probably using a beam antenna of some sort, and, as often as not, using the same beam antenna for receiving and transmitting. If you hear a tasty DX station arriving from afar, your response is to steer your receiving antenna for maximum received signal strength, with the assumption that this will orient your antenna for the best signal at *his* location as well. As the previous discussion strongly implies, this is a *bad* assumption! The magnetic field almost certainly assures that the path is *not* reciprocal…and it's certainly not a Great Circle path.

If you conclude that it would be beneficial to have separate transmitting and receiving antennas, you would be absolutely correct! Now the question remains is this: how do you know where to steer your transmitting beam if you can't rely on your receiving antenna's heading? This most likely would require that you actually *ask* the distant station how your signal is. It might take a bit of give and take to get your antenna aimed down his boresight…and the same for him on his end…assuming he also has two antennas!

The real issue is that to really take advantage of the way radio *really* propagates, we need some sort of feedback. Verbal feedback is fine; it gives you something to talk about while homing in your antennas! But, more interestingly, it suggests some interesting possibilities for closed loop telemetry tricks as well. (See the sidebar, "Taking Advantage of X and O Modes.")

This operating practice might also promote the use of honest signal reports, as well.

Now, while this may seem somewhat tedious (and slow) for contest operation, for DX ragchewing, this could greatly increase your window of opportunity. Standard practice is to assume if you're experiencing fading on the DX station's signal, so is he or she, and thus it's a hint that you should be getting ready to end the contact soon. But once it's understood that propagation paths (and current conditions) are not reciprocal at all, you can adjust your operation accordingly. One can never know how many QSOs were unnecessarily lost because of not understanding the "one way" nature of most HF conditions.

Taking Advantage of CPOL

There probably aren't more than a handful of hams actually using circular polarization on HF As it turns out, without at least a CPOL receiving antenna, you can never know for sure whether you're receiving the distant station's X or O mode signal. You can sometimes *guess*, based on the fact that O mode signals are generally stronger, and more closely follow great circle paths. But this is only an educated guess. The fact of the matter is, to take advantage of X and O mode propagation, you need to be circularly polarized, just as the signals themselves are.

Now, although a linearly polarized receiving antenna responds just as well to a circularly polarized signal, regardless of the polarization, there will always be a 3 dB loss. This means that just by "going CPOL" will gain you 3 dB of signal strength over a linearly polarized antenna of the same gain. This is, of course, assuming that your CPOL antenna is the right *sense*. If your receiving antenna is the *wrong* sense, it can reduce the signal to nothing…at least in theory.

As it turns out, under typical conditions, a hastily constructed CPOL receiving antenna will show about a 3 S unit difference between the "correct" sense and the opposite one on *any* HF ionospherically propagated signal. This is pretty compelling evidence for the existence of X and O modes. Most hams are astonished when they perform this simple experiment. A simple, practical antenna for experimenting with HF CPOL is a crossed inverted V antenna with a simple 90 degree phasing stub. It requires only a single support pole, and has a reasonably small footprint even on 40 meters, where CPOL effects are most evident. (Again, see the sidebar, "Taking Advantage of X and O Modes.")

What About X?

Obviously, radio amateurs have done just fine for decades without knowing anything about the X-mode ray…or the O-mode ray. Again, it's a true testimony to the romance of Amateur Radio that we can do so much while knowing so little! But now the cat is out of the bag, and it will be pretty hard to stuff it back in. Hams now know about X and O mode propagation, and once something is known it can't be unknown.

It's been suggested that, by default, most hams on a daily basis deal mainly with the O-mode signal, mainly because it tends to be stronger, and a little more "conventional" in behavior. But the fact is that *every time* you launch an HF signal into the ionosphere, you generate both an X and an O ray. Approximately half of all Amateur Radio power transmitted in the world is in an X-mode ray….somewhere. And that is the big "X Factor." Where does that X-mode ray go? Where on Earth *is* half of my radiated power ending up?

As we stated earlier, you can't know for sure without a CPOL antenna. It doesn't mean you can't receive an X-mode signal just fine with a dipole…or a vertical…or a Yagi…or any other kind of antenna. It's just that you don't know for sure if it's an X-mode signal or not. But we can certainly use an educated guess even if we don't happen to have a CPOL receiving antenna. We know that in the northern hemisphere (where the vast majority of hams reside anyway), the X mode ray is going to diverge toward the magnetic equator from where it's "supposed" to go. So, this in itself can give you a strong hint. If I'm in Hawaii, and I'm hearing a station in California who has the antenna aimed at Alaska, it's a fair bet I'm receiving the "X-ray." By the same token, if I'm in Fairbanks, and I hear a Californian who swears he's aimed at Hawaii, it's almost certainly I'm receiving his O-mode signal.

Again, as demonstrated in Chapter 6, this effect depends on frequency. In the example in Figure 6.4, there is little deviation in azimuth by either the O-wave and X-wave on 10 meters. More investigation (or modeling!) is warranted.

Taking Advantage of X and O Modes

The material presented here is an *extremely* abbreviated introduction to a long and complex topic, but one that is inseparable from a topic that is dear to most radio amateurs: HF radio propagation. The following topic might be considered highly controversial if it weren't for the fact that it's been barely mentioned at all. A controversy, by definition, requires two opposing voices. In this instance; one voice has been almost entirely absent!

For many decades, a lot of unexplained "bizarre" radio behavior has been summarily dismissed as being…well…just bizarre unexplained radio behavior. Part of the charm and romance of Amateur Radio is, of course, due to the fact that we don't know everything, and it is always a thrill to discover something that nobody else has observed. However, a good deal of what we simply write off as inexplicable is indeed fully the result of very well-known physical principles. Real progress is made when we logically and systematically study our observations, no matter how bizarre, erratic, or random they may seem.

Many of us have heard it said that HF radio signals are "randomly" polarized. This not only does a great disservice to rigorous radio science practitioners, but to the physical universe itself, which is a *whole lot* more consistent than most of us give it credit for. As it turns out, with extremely few exceptions, most of what we observe with regard to HF radio propagation can be explained using known radio physics. It only becomes "random" when we ignore one of the main ingredients of radio propagation: the Earth's ever-present magnetic field. When you ignore it, almost nothing makes sense. When you include it, nearly everything makes sense.

The fact that we have been able to explain radio propagation with *any* consistency at all, while omitting such a crucial factor, is astonishing in itself.

Many old timers will argue till they're blue in the face that there is "no such thing" as one way propagation. This totally misapplied "law of reciprocity" has been *propagated*, pardon the pun, so consistently and so effectively, and for so long, that it has taken on the totally undeserved fragrance of truth, despite all empirical evidence...and physical law….to the contrary. When we understand the birefringent nature of a magnetized plasma, the fact that we *ever* have anything resembling reciprocal propagation is nearly miraculous! As we will discover, it is only due to the fact that the average HF antenna is so *bad* that we can use the same for both transmitting and receiving over the typical long distance HF path! It's a good thing that most of us are using high gain Yagis at the most, and not lasers; in the latter case the chances of both paths crossing each other would be nearly non-existent!

Another factor very few hams consider is the *incoming vertical angle of arrival* of distant signals, which is actually a lot more important than is typically acknowledged. We very seldom notice this, however, because in order to accurately ascertain the direction of arrival of a signal, your antenna needs to match the *polarization* of the incoming signal. Since, as we will soon see, HF ionospherically propagated signals are, without exception, *circularly polarized*¸ virtually no radio amateurs are capable of accurately determining this direction of arrival — much less take advantage of the unique properties this circular polarization has to offer.

In this very short discussion, we cannot delve very far into the inner sanctum of the physics that causes these unique and fascinating circularly polarized waves to emerge. However, we *can* treat the magnetized ionosphere as a "black box" with extremely predictable "output" behavior based on certain "input" behavior.

When a linearly polarized radio wave enters the ionosphere, the interaction of the Earth's magnetic field and the ions in the path causes the radio wave to split into two separate rays, one being clockwise circularly polarized, and the other being counterclockwise circularly polarized. This transition does not happen instantaneously, but rather gradually, as the radio passes through a medium of *continually* changing density gradient. These two separate rays, known as the X-mode ray and the O-mode rays, diverge from one another in azimuth, vertical angle, and velocity. The amount of attenuation depends on the direction these rays travel with respect to the magnetic fields.

These paths are neither reciprocal nor do they follow Great Circle paths….except for statistically rare occurrences. The only case where the X and O rays *would* follow nearly identical paths would be of both the transmitting and receiving stations were precisely on the magnetic equator, which would be about as likely as a Congressman voting himself a pay cut.

The result of all this is not only fascinating, but also, ultimately, quite *useful.* If we were to deliberately *select* one or the other of these rays, we could achieve frequency re-use of nearly *any* HF signal path, amongst other things.

Knowing that the X and the O mode signals have different *reflection heights* should present some obvious benefits. We can select our skip distance by selecting our polarization, for example. There are a few complications here.

One of the "odder" properties of X and O propagation is this: If you *launch* a linear polarized signal, it splits into the X and O modes as we demonstrated above. If, on the other hand, you *launch* an X or O mode signal (by means of an appropriately polarized CPOL transmitting antenna), it tends to *stay* an X or an O mode signal…whichever way you launched it. The physics behind this is well beyond the scope of this book but is a field ripe for amateur experimentation!

From a more practical standpoint, the average ham will benefit more from the use of a CPOL *receiving* antenna, generally speaking. In addition, most hams need to *prove* it to themselves that X and O modes actually exist in the first place before they invest a lot of trouble and expense into converting their entire stations into X and O mode juggernauts. It is extremely easy to build a CPOL receiving antenna to demonstrate the profound circularity of HF propagated signals. Even a hastily assembled CPOL turnstile will show a 3 S unit difference in strength between "correct" circular polarization sense and "wrong" circularly polarized sense. Most hams are absolutely astonished to discover this, even those with many years of on the air operation and experience with HF propagation.

Once you have seen for yourself that HF signals are indeed circularly polarized, it's usually a small step to the "so what" of all this. This knowledge itself suggests all kinds of new experiments and also suggests new ways of "doing business" as HF radio operators.

This introduction is by no means the last word on the topic. In most cases it is the *first word* on the topic on X and O propagation for the vast majority of hams. I trust a number of you will test this out and see for yourself! Don't take my word for it. All scientific progress requires a healthy gift of suspicion.

At the very least, I hope that this new knowledge helps explain what you have already observed, and goes a way toward removing some of the "chaos theory" you have been subjected to in your ham radio operation. — *Eric Nichols, KL7AJ*

Forgiven

Most hams can be forgiven for not knowing anything about the circular polarized nature of HF signals (at least before they read this book). CPOL signals can be received on linearly polarized antennas just fine, and unless you're actually *looking* for such things as X and O rays, you can go a long time without noticing them! In fact, if it weren't for the optics folks, we might still have no clue about the characteristic rays of the ionosphere. Ionospheric physicists discovered X and O mode propagation because they were looking for it! The principles of *birefringence* were known long before from experiments with crystals and magnets and polarized candle light.

Not So Random

For many decades, the "wisdom" in Amateur Radio was that the polarization of HF antennas didn't matter too much because HF reflected signals were "randomly" polarized anyway. While it is true that, statistically speaking, a randomly polarized signal can be received some of the time on a randomly polarized antenna, the actual results are generally much better than this can account for.

The fact of the matter is that vertical or horizontal antennas work pretty doggone well most of the time for HF signals of any origin. This can only be true if the signals are essentially circularly polarized, which an antenna of any polarization can respond to well. Again, we may have a 3 dB loss over a properly polarized CPOL antenna, but a lot of other factors can account for a lot more loss than that. It is indeed the fact that X and O modes exist that makes Amateur Radio as *forgiving* as it is!

Getting Flighty

One tool that the professional HF scientists have that we amateurs don't (at least directly) is the ionosonde. Because an ionosonde is a radar, it can derive *time-of-flight* information that is not readily available to "normal" Amateur Radio methodology. Because the virtual heights of X and O mode reflections are significantly different, they show up very clearly as double traces on an ionosonde, even *without* the use of a CPOL receiving antennas. Early scientific researchers saw these double traces and immediately suspected X and O modes…because such things were already on their "radar"! All it took was a CPOL receiving antenna to confirm their suspicion.

Both

A vertical incident ionosonde is one of the few cases where you'll see the X and O mode rays at the same place. In such an instance, they haven't traveled far enough to diverge very far, and since the receiving antennas are pretty wide-beamed, they can still "see" the X and O modes from essentially the same place. This is *not* likely to happen for typical amateur DX operation, though for NVIS operation (quite common in Interior Alaska, for instance), both the X and O modes can be available at the same receiver. So the obvious question is, what happens if an X and O mode signal are picked up by the same antenna at the same time?

On the practical side, this is never going to happen. As we've just discussed, the time of flight of the X and O signals is different, so even if you receive them both on the same antenna, you won't receive them at the same time. In cases where you do receive them both, it can sound like a very short-term echo…actually a rather common occurrence for NVIS operation. The typical "hollow, reverb-y" sound of many NVIS signals are due to the slight time delay between the X and O mode signal arrivals — generally on the order of a few milliseconds.

On the purely theoretical, ideal side of this, here's what happens. If an X and an O signal of equal intensity arrive at precisely the same time on a *linearly polarized antenna*, they will precisely cancel out; you will get nothing. On the other hand, if you are using a CPOL antenna, you will only receive the ray for which your antenna is matched (of the same *sense*).

You Send Me

Here's something to ruminate on. We have described how any linearly polarized signal becomes circularly polarized upon entering the ionosphere, splitting into both an X and O mode ray. What happens if we *start out* with a circularly polarized signal?

As it turns out, under *most* cases, the signal will remain whatever sense it started out in. If you transmit an X mode signal, it tends to stay X mode throughout its journey. If you transmit an O mode signal, it tends to stay that way too. We can already hear the collective gears grinding within thousands of amateurs' heads as we mention this. Does this mean we can use polarization to actually *steer* our signals? Absolutely! Now, it should be evident that you don't have quite as many options here. A signal can be either X or O, it can't be much of anything in between. So if you launch an X signal, it will follow the normal X path, if you send an O signal, it will follow the normal O path (for a given set of conditions, of course). Are we having fun yet?

Dealing with Reversals

We'll conclude with a phenomenon that actually has nothing to do with propagation, but with some simple…well, moderately so….geometry. This is something you need to consider if you plan on going whole hog with circular polarization on HF

When you transmit a right hand CPOL signal through free space, the antenna at the receiving end has to be *left hand* CPOL. (VHF folks know all about this.) You can easily demonstrate this by taking two wood screws and aiming them at each other point-to-point. Which way do you have to turn the screw to make it go forward? Which way do you have to turn it to make it come *toward* you…as in the case of an incoming signal? What this means is that if you're set up for CPOL, you either want to use a separate receiving antenna of the correct sense (the ideal case), or have a sense reversing relay tied to your push to talk circuitry, in the case of a single antenna for transmit and receive.

Try this at Home

This chapter on X and O modes is certainly not the last word on the topic. It is our sincere desire that we have encouraged you to experiment with these fascinating modes. It's one of the areas available for us to make meaningful strides toward "advancing the state of the radio art," which is what we're all called to do in some capacity.

Instrumentation and Interpretation

Whether you plan on just using the ionosphere, or wish to make a meaningful contribution to understanding the ionosphere, there are a number of instruments and practices at your disposal. Every radio amateur should know how to read an ionogram. Although sophisticated instruments such as the Lowell Digisonde present a lot more information than you really need, the general principles should be well understood. The ionosonde is probably the most useful instrument for both predicting radio propagation and for understanding the conditions that currently exist.

For a great overview of ionospheric sounding, here is the place to start: **www.digisonde.com/instrument-description.html**

As we mentioned in the previous chapter, hams don't have direct access to radar methods, but Digisonde data is widely available online, as are data from countless other instruments around the world.

Keep in mind that the world is big, *very* big. Someone else's ionosphere is not *your* ionosphere. You want to get good data not only for your location but for the location of the DX station you want to work, and as many points as you can in between. Most hams agree this is preferable to using random dumb luck…though not always as romantic. It should be noted here that, as complex and vast as the ionosphere is, it's unlikely that we will know so much about it that the mystery, romance, and intrigue will be lost. It is safe to say that will never happen. Chances are, the more you know about the ionosphere (or any other topic of scientific interest) the more awestruck you will be at the wonder of our natural universe.

Being Scientific About It

Amateur Radio is a scientific hobby, and it's always a good idea to review the principles of good scientific method; even we old timers need reminders from time to time. Maybe *especially* we old timers! Here are just a few of those principles which will make your Amateur Radio technological progress more fruitful.

- *The physical universe is governed by laws, not suggestions.* We don't know everything, but this does not mean things are not knowable. Scientific laws exist with or without our help. Our job is to discover those laws, and make use of them where appropriate.
- *Technology is not science*. Although we hams make use of both science and technology, there are important differences. Technology is how we apply scientific principles to practical matters. This is an art, and art, like every other human endeavor, is subject to change. Science is based on unchangeable laws.
- *Scientific law is universal…and technology nearly so*. A Russian schematic diagram looks pretty much like an American schematic diagram. Mathematics supersedes politics. There's a reason we amateurs are in a unique position to "promote international good will" as spelled out in our *raison d'etre* in FCC Part 97.1e.
- *Good science is subject to peer review*. If something sounds fishy, it probably is. There are snake

oil salesmen in Amateur Radio, as in every other endeavor, but collectively we should be able to sniff them out.

• *Somebody has probably already thought of it before…but don't count on it*. Hams will continue to make new discoveries as long as we exist.

• *The best thing that could happen to a ham is to have a scientific law named after him or her*. While this is not absolute scientific truth, it probably passes the peer review test.

With these principles in mind, let us now look at how we can take a more scientific approach to our Amateur Radio operation and activities. For the relative newcomer, we trust that you recognize your ham ticket as merely the entry point, not the destination, of Amateur Radio knowledge.

There is no better tool for learning about radio propagation than a good old fashioned shortwave receiver, now more commonly known as a general coverage receiver. Some truly excellent receivers are available for a song, but it's a good idea to get the best receiver you can afford. A "dc to daylight" communications receiver, which includes VLF, LF, MF, HF, VHF, UHF, and perhaps a microwave band or two, is certainly going to be more of an investment than a classic boat anchor, but will be a lot more useful instrument in the long run.

Beacons Beckoning

You should be aware of WWV, WWVH, and other time and frequency stations. There is a complete description of these valuable stations in the *ARRL Handbook* and elsewhere. WWV, because it broadcasts on numerous HF bands throughout the shortwave spectrum, serves as a reliable general propagation beacon. While it won't tell you much about what conditions are in Asia or Africa, for the most part it will tell you the general state of health of the ionosphere. Beyond WWV, we have countless worldwide shortwave broadcast stations. Despite the predictions that shortwave broadcasting was going to go away (and for a short while, it almost seemed this was the case), there has been a resurgence of activity throughout the shortwave bands — both friendly and nefarious — that gives us assurance that, while the players may change, there will always be players. Shortwave bands are alive and kicking! Just give a listen!

By the way, since WWV sends "ticks" along with all the other information, you can actually get time-of-flight information. You can also sometimes hear the time lag between the X and O signals from WWV as a subtle double click where there should be none. (WWV does send periodic double clicks. Again, refer to the WWV schedule.)

The S Meter is Your Friend

One of the most annoying habits of modern "hamdom" is the general disregard for honest signal strength reporting. Traditionally, signal reporting was one of the most important activities in Amateur Radio. The development of an actual functional S meter (**Figure 9.1**) was a major innovation at the time, and its advent should not be taken lightly! Many hams dismiss S meter readings as being "meaningless," since what you're really looking for is signal-to-noise ratio (S/N), anyway. Like so many aspects of our hobby, a little knowledge is a dangerous thing. Certainly signal-to-noise is important for communications intelligibility, but it's not the only factor…especially when it comes to actually making advances in the radio art.

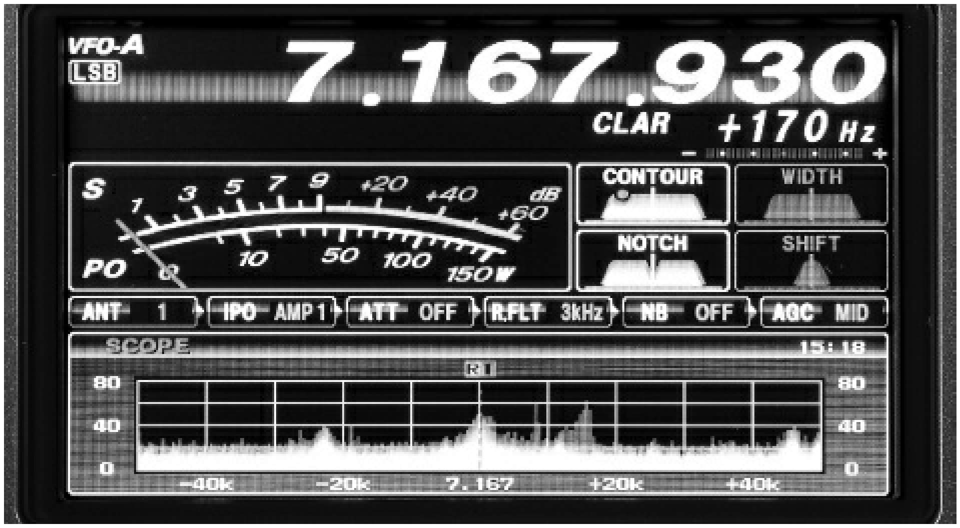


Figure 9.1 — In modern transceivers, the S meter is often "virtual" — an on-screen representation rather than an actual meter.

A communications receiver with a well-calibrated S meter is an extremely valuable instrument. A lot of smart folks collaborated in the design of the standardized S meter format, and it hasn't been significantly improved upon in many decades. In addition, we now have at our disposal extremely inexpensive and accurate data acquisition equipment, which can make recording signal strengths very simple (and a lot of fun).

We'll discuss S meters thoroughly in Chapter 17, Your Friend the S Meter.

Channel Probes

Most radio amateurs are familiar with the many Amateur Radio beacons, especially on 10 and 6 meters. These are great for learning when there's a band opening, and from where. But we can greatly expand the usefulness of beacon stations by means of the *channel probe.* The channel probe is simply a beacon station with a dedicated monitor station at some distant location. Channel probes can be quite elaborate, able to resolve polarization angles, angle of arrival (often in two axes), and sometimes even total time of flight.

Channel probes have been around a long time in the professional ionospheric circles, but are just coming into their own in Amateur Radio, most frequently in the form of "grabber stations." Grabber stations are remote-controlled receivers accessible via the Internet. A grabber station, among other things, allows you to make real time corrections to your transmitting setup to achieve the best signal at the target location.

You can experience channel probes yourself with the Reverse Beacon Network (**www.reversebeacon.net**), a network of "skimmer" stations around the world. The skimmer stations receive CW and RTTY signals across the amateur spectrum and report back on the Internet with call signs heard and signal strengths received (**Figure 9.2**). If you're wondering if a band is open, or where it might be open to, simply send a few CQs and monitor the RBN to see if or where your signal is heard, and how strong it is.

This idea will become more and more important as more and more hams adopt X and O communications methods as described above. Virtually any ham radio station can become a grabber station, if an Internet connection is available. But there are also "older school" telemetry methods available for sending S meter readings and other information back to the transmitting station "in band."

band: 160m,80m,60m,40m,30m,20m,17m,15m,12m,10m   rows to show: 15 ▼

cancel filter selection / search spot by callsign

de	dx	freq	cq/dx	snr	speed	time
OH6BG	YU3AO	7031.5	CW CQ	12 dB	22 wpm	1541z 14 Jan
OL5Q	YU3AO	7031.5	CW CQ	23 dB	23 wpm	1541z 14 Jan
SK3W	YU3AO	7031.5	CW CQ	20 dB	22 wpm	1541z 14 Jan
DL8LAS	YU3AO	7031.5	CW CQ	21 dB	22 wpm	1541z 14 Jan
DL8LAS	OK1IF	7023.1	CW CQ	17 dB	20 wpm	1541z 14 Jan
KS4XQ	KM6JD	28026.7	CW CQ [LoTW]	23 dB	22 wpm	1541z 14 Jan
AA4VV	KM6JD	28026.6	CW CQ [LoTW]	18 dB	22 wpm	1541z 14 Jan
N7TR	HC2IMP	28014.0	CW CQ [LoTW]	6 dB	23 wpm	1541z 14 Jan
N8MSA	KM6JD	28026.6	CW CQ [LoTW]	36 dB	22 wpm	1541z 14 Jan
DF4UE	YU3AO	7031.5	CW CQ	14 dB	22 wpm	1541z 14 Jan
DF4UE	VE9NC	24925.43	RTTY CQ [LoTW]	23 dB	45 bps	1541z 14 Jan
W8WWV	IZ5FDE	21013.4	CW CQ	7 dB	27 wpm	1541z 14 Jan
W8WWV	K3AM	21033.6	CW CQ	11 dB	14 wpm	1541z 14 Jan
K8ND	KM6JD	28026.6	CW CQ [LoTW]	35 dB	22 wpm	1541z 14 Jan
ON5KQ	OH2B	18110.1	CW NCDXF	26 dB	22 wpm	1541z 14 Jan

Figure 9.2 — The Reverse Beacon Network reports (left to right) call sign of the receiving "skimmer" station, call sign of the station heard, frequency, mode, signal-to-noise ration of the received signal, sending speed, time and date. In this case the received signals are mostly CW, along with one FSK RTTY signal and a Northern California DX Foundation (NCDXF) propagation beacon.

Sound Advice

Let us now return to the venerable Digisonde and see what else we can learn. Eielson AFB near Fairbanks, Alaska, recently deployed a Digisonde, much to our delight. We've been without an ionosonde in this area for many years, and it's great to see a sounder online again.

In **Figure 9.3** above we have a fabulous example of sporadic E as seen on an ionogram. Notice the perfectly horizontal trace at about 110 km height. We know this is sporadic E because it is frequency independent. But notice the three other horizontal traces at 220 km, 330 km and 440 km. These are actually multiple hops…straight up and down. This tells us that the absorption is *extremely* low. As noted in Chapter 6, in the discussion of Figure 6.6, what's really happening is the energy is going up, coming back to Earth, going back up and then coming back down. Since we're really measuring time of flight, it shows up as 2, 3 and 4 times the original altitude since it's 2, 3 and 4 times the time of flight.

These are perfect conditions for some smoking hot high-band DX. Notice also that we see normal F2 traces above the sporadic E, as well (again with multiple reflections). This reveals something very interesting — the fact that as *specular* as sporadic E reflections are, they are not *total*. Enough RF gets past the E layer to undergo a nice reflection at the F2 layer too. This is not always the case; often when there is a strong sporadic E event, you don't see anything happening above it. This gives further evidence that the ionization process for F2 is very different than that for sporadic E, which is most likely activated by highly localized sources, such as lightning, wind shear, or meteor bursts. These events are completely independent of "normal" solar radiation ionization.

Learning how to interpret ionograms completely takes a little practice, but it's a lot of fun too. You are encouraged to look at whatever Digisonde is closest to you to see if you can correlate its data to the actual propagation you're experiencing. Check out the Digisonde station list at www.digisonde.com/stationlist.html. It shows every Digisonde in the world that's available online (and a few that aren't).

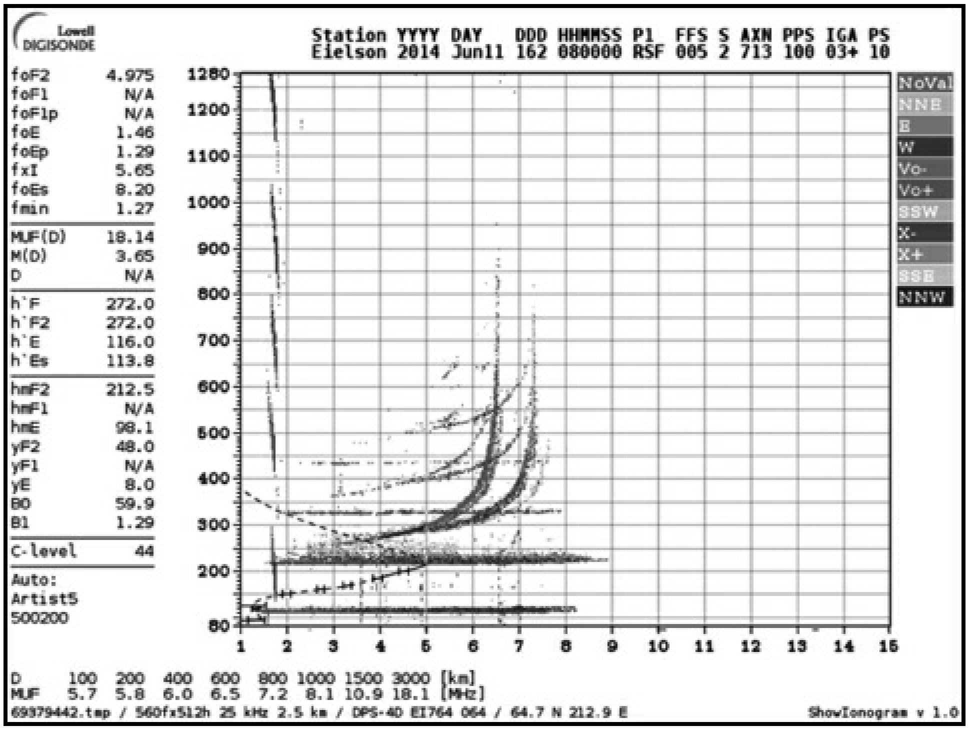


Figure 9.3 — Ionogram showing sporadic E propagation. See text for details.

A few of the sites may need a password to log in. Most of these sites are pretty accommodating to radio amateurs; if you just drop the webmaster a note, they'll often give you a temporary login. As a general rule, Digisondes are programmed to perform a sounding every fifteen minutes; and you can usually access archived data back at least a couple of years. Some sounders are set up to perform oblique (non-vertical incident) sounding as well, so be sure to read all the supplementary data that's on the ionogram.

A great reference is UAG-23A, the *URSI Handbook of Ionogram Interpretation and Reduction*. This publication contains the agreed standard rules for interpreting ionograms. You can find the latest edition online at www.ursi.org/files/CommissionWebsites/INAG/index.html.

Dabbling in Doppler

The ionosphere is not chiseled in granite; it moves around a lot. Most Digisondes are set up to show *Doppler shift* resulting from the movement of the reflection point. Doppler shift is shown by a shift in the color of the traces. For the O-mode, downward movement will show a shift toward purple, and upward movement will show a shift toward orange, with monochromatic red being stationary. (This is equivalent to Red and Blue shift, for the astronomers in our midst).

As mentioned earlier, Doppler shift can be extreme during nightfall, but there will almost always be some present even during the most stable conditions. If the Digisonde is set up for oblique sounding, Doppler shift will show drift of the ionosphere in azimuth; in effect it shows ionospheric "wind."

Again, be sure the check all the entries in the legend of the particular ionogram in question. Most of the time there will be quite a few blank entries, but if the ionosonde is set up with all the possible features, the entire column on the left may be full of meaningful data!

Other Online Instruments

For the average ham, next to the Digisonde, magnetometer data is probably the most useful data, as it can actually *predict* certain aspects of propagation. Like the Digisonde, online magnetometers abound, but you may have to look a bit harder for them. While the Digisonde pretty much dominates the ionosonde market there are countless magnetometer manufacturers. Many of them are semi-commercial homebrew instruments, being constructed by university students and such, so you won't find a single magnetometer station list (as of this time).

Here is a typical magnetometer website, this one being from the Swedish Institute of Space Physics: **www.irf.se**. Click on Observatory and then Magnetometers.

This displays the data from a typical three-axis magnetometer, X being north-south, Y being east-west, and Z being up-down. There's a lot of useful and educational supplemental information on this site, too. It is common to use a *flux gate* magnetometer to detect the Earth's magnetic field...actually the *change* in the Earth's magnetic field. Magnetometer readings are *normalized* to calm day magnetism. If you look at the magnetometer plots in the example, you will see a rather sharp spike on a couple of the axes, which indicates a pretty strong disturbance. Being a three-axis device, we can know not only the strength, but the *direction* of the disturbance in three dimensions.

The much-obsessed-over *solar indices* can be derived from local magnetometer data. More on these factors will be coming up later.

Another instrument that is of particular use in Polar regions is known as the *riometer*, for *relative ionosphere opacity meter*. This device is a sensitive receiver operating in the 30 MHz region that looks for diurnal variations in cosmic noise, and is an indicator of ionospheric *absorption.* Auroral activity is generally highly absorptive, and auroral events are clearly seen on the riometer. However, ionospheric opacity is a variable that can have large effects in nonpolar regions, as well.

Index Finger

We now come to the most common, if not the most reliable, predictors of ionospheric propagation, the solar indices. There are three primary factors that predict what propagation on Earth will be like: the solar flux, the K index and the A index. Although these are important factors, they are not the only factors, and they need to be taken with a certain degree of educated interpretation, if not a grain of salt or two.

Before going into these too far, we need to recognize a few things. Number one, the ionosphere is a lot closer to us than the Sun! The Sun is about 93,000,000 miles away, while the ionosphere is between 60 and 400 miles away, at the far edges. Solar indices can tell you what *may* happen, while the ionosphere tells you what *is happening.*

Just as weather "hindcasting" is always more accurate than weather forecasting, ionospheric sounding is always more accurate than ionospheric prediction. This doesn't mean that it's always more *useful*. If you're a contester, you'd like to know what conditions are going to be during the upcoming weekend, to know if it's worth bothering to turn on the rig! And if conditions are currently *good*, you probably don't need an ionosonde to tell you that! So we need to know what we're looking for, and adjust our prediction and measurement methods accordingly.

Suntan

We described earlier how the best condition for creating a consistent ionosphere is a steady bath of UV radiation. The solar flux is measured in solar flux units (SFU), where

$$SFU = 10^{-22}\ W\ m^{-2}\ Hz^{-1}$$

is a numerical indicator of the amount of UV radiation coming from the Sun. It is, by tradition, (and nothing else!) measured at a wavelengths of 10.7 cm. The figure is a measurement of the solar radio noise detected at 2.8 GHz (originally 2.695 GHz). The *official* measuring station for this figure is the Penticton Radio Observatory in British Columbia, Canada, but if you like, you can fairly easily build your own 10.7 cm monitoring station. The solar flux is a good indication of the overall *electromagnetic radiation* hitting the Earth, and thus a good indicator of the overall ionization level of the ionosphere. There is a fairly direct relationship between sunspot numbers and solar flux, expressed in the simplified formula:

$$SF = 73.4 + 0.62\ R$$

where SF is solar flux and R is the daily sunspot count. A slightly more accurate formula is:

$$SF = 63.7 + 0.73\ R + 0.0009\ R^2$$

where again R is the daily sunspot count.

Note that these equations are for *smoothed* sunspot number and smoothed solar flux. The correlation between the *daily* sunspot number and daily solar flux is poor.

If you are so inclined, you can count sunspots yourself, (being sure to use proper safety procedures, of course!) and compare your calculations to the periodic broadcasts of solar flux by WWV.

Now, since solar flux is an indicator of electromagnetic radiation, which travels at the speed of light, it is a fairly immediate indication of solar UV radiation. It only takes eight minutes for electromagnetic radiation to get from the Sun to us. However, the ionization process is not instantaneous, so we generally need to look at a day or so of solar flux numbers to get meaningful propagation information. In addition, sunspots themselves form and disappear over the time frame of several hours at the fastest. The SFU, by the way, can run as high as 300. In general, high SFU numbers are beneficial, but they aren't the only story.

Solar UV radiation isn't the only factor that determines the state of the ionosphere at any given location. Geomagnetic field activity and events in the lower atmosphere also contribute to the ultimate ionization — and this is why we have a monthly median ionosphere model.

Particular about Particles

Although we like a strong bath of UV radiation, we don't particularly want *high energy particles*. Whenever there is a lot of sunspot activity, there is the potential for fast moving electrons and protons to be ejected. These particles, since they are charge carriers, both create and disrupt magnetic fields when they come in close proximity to the Earth. In particular, when these particles *cross* magnetic field lines, they become part of an electrical generator.

The aforementioned magnetometers can measure the contribution of these charged particles to the Earth's field. Two indices, the A index and the K index are measured by the horizontal magnetometers (the X and Y devices). The Z axis magnetometer is primarily unaffected by particles coming in perpendicular to the Earth's surface (and thus parallel to the Z magnetometer's magnetic field), and are therefore not part of the K and A index measurements.

Charged particles are generally disruptive to propagation, since they disturb the smooth layering of the ionosphere, among other things. The best conditions exist, therefore, when we have a high solar flux, (again, resulting from strong electromagnetic radiation), along with low magnetic activity (a result of minimal charged particles arriving).

Now, since charged particles have real mass, they take a lot longer to get to Earth than the UV radiation, on time scales ranging from hours or days to even weeks. Hold that thought.

What's in a Number?

While the SFU is an actual unit (defined above), the magnetic indices are semi-arbitrary scales derived from magnetic information, and thus are referred to as *indices*…or indicators of some characteristic. Let's look at the K index first.

There is a network of magnetometers scattered around the Earth. They are all calibrated so that a given level of magnetic disturbance gives the same indication at each site. This is why it's more specifically called the *planetary* K index or K_p. The K index is registered on a scale of 0 –9, with 0 being calm and 9 being highly disturbed. Now, it probably goes without saying that the disturbance of the magnetic field is highly dependent on location, which is why all the sites are calibrated to a known standard. The actual reading is also logarithmic, which means direct averaging of various sites is impractical. As useful as the K index is, it's a little bit *unwieldy*, for lack of a better term. So the A index was developed. The A index is measured with the same instruments, but averaged on a daily basis. It is also scaled differently. **Table 9.1** shows the relationship between the A and K scales.

Geomagnetic storms are seldom beneficial for radio propagation. While there is a connection between geomagnetic storms and ionospheric storms, there can be a time lag between them as well. Again, ideal conditions would preclude any high geomagnetic numbers of either the A or K sort. But often, some unusual propagation can exist in the midst of a geomagnetic storm; a storm doesn't necessarily mean a radio blackout.

We should also mention that the A index varies from station to station; while it represents an average over *time*, it is not normalized with respect to location as is the K index. This means that both the A and K indices have their places, and it's probably not best practice to look only at the K index, as is sometimes advised in recent years.

The Whole Elephant

We've all heard the allegory of the three blind men attempting to describe an elephant. One blind man takes hold of a leg and declares, "The elephant is like a tree." The second blind man grabs an ear and says, "The elephant is like a big leaf." The third one finds its trunk and says, "The elephant is like a fire hose."

Table 9.1
Numerical Relationship between A and K Values

A	*K*	*Condition*
0	0	Quiet
2	1	Quiet
3	1	Quiet
4	1	Quiet to unsettled
7	2	Unsettled
15	3	Active
27	4	Active
48	5	Minor storm
80	6	Major storm
132	7	Severe storm
208	8	Very major storm
400	9	Very major storm

All these are apt descriptions, but fail to describe an elephant to a fourth party adequately. Attempting to predict future radio propagation by means of A and K indices, which are *already* rather indirect indications of solar activity, is unpredictable, at best. Since particle activity can lag electromagnetic activity by hours or days, even having all three ingredients of the elephant — the solar flux, the A index, and the K index — may not be enough, since they don't even occur at the same time!

Propagation is an art form, and while we have been getting somewhat better at it over the past century, it is far from a hard science. We still can't predict when…or how many…sunspots may pop up at any time. And we are *still* not sure where we are in the current sunspot cycle. Top experts still can't agree on whether we're going to have the best or the worst solar cycle in the history of ham radio…or the history of Man! It is this very unpredictability that helps make Amateur Radio so much *fun*!

Your Personal Scientific Toolbox

While it is important to know how to collect and interpret data that other folks have generated, there are a lot of interesting measurements you can make yourself. In fact, you can make a lot of the instruments yourself. Here are just a few suggestions:

1) *A field strength meter.* This doesn't need to be precise, or calibrated to an absolute standard for a lot of radio work. On the other hand, an absolute field strength meter can be valuable for doing advanced propagation studies. Field strength is measured in volts per meter (V/m). A 1 meter long piece of wire "dipped" in a field of 1 volt per meter will have a 1 V difference of potential between its ends. While extremely simple in concept, it's a bit trickier to actually measure the voltage across the ends of such a wire. Most of us are used to measuring voltage at the *middle* of a wire! Once you break the wire to insert a voltmeter, you no longer have a 1 meter piece of wire. If you leave the wire intact, but try to measure the voltage at the ends, your voltage meter leads are going to be part of the antenna and mess things up too! So, an *actual* calibrated field strength meter is generally built around some form of *loop* antenna which is calibrated to be equivalent to a straight piece of wire. The math to do this is beyond the scope of this book, which is why such instruments can be a bit pricey. However, if you have a *known* field strength, you can use that as a calibration standard for a much simpler meter. The *ARRL Antenna Book* has a field strength meter that can be calibrated to an

absolute standard in this manner.

2) *A reference transmitter.* This is a transmitter of known power, feeding an antenna of known gain. While a dipole is the easiest thing to calibrate in this regard, there's nothing that says you can't use an antenna with significant "negative gain." This is useful for calibrating your field strength meter (described above), as well as providing a source for general antenna gain measurements.

3) *A good DF (direction finding) antenna (or antennas) for your frequency of interest*. At lower frequencies, a multiturn tuned loop, preferably shielded, is pretty much standard. Such a loop provides a sharp null in the direction of the radio source. At VHF and UHF, standard practice is to use a high gain Yagi and look for the maximum signal strength. This method is seldom as accurate as a nulling method, however. Various forms of interferometers are well suited to VHF direction finding. Being able to direction find, both in the azimuth *and* in the vertical plane, is a crucial part of a lot of ionospheric study.

4) *A good general coverage receiver*. This is probably the best test tool for *any* ham radio shack, but for propagation studies it is crucial. You want the receiver to be as stable as humanly possible (or affordable). You don't need a lot of bells and whistles, but you want a clean receiver, with a minimum of "birdies" and other weird artifacts. One of the best "science radios" is the direct conversion receiver, an item that's easy to build for the average ham.

5) *A computer with a sound card and "waterfall" software*. Any of the HF digital software packages work fine. It doesn't need to be a screaming fast machine, though it helps to have some horsepower if you're doing propagation modeling with a program such as *Proplab Pro*.

6) *An oscilloscope*. This doesn't have to be fancy either. Some of the best propagation experiments are "self-correcting," meaning you can use some pretty primitive equipment to get wonderful results.

7) *Another ham across the country*. This may be your most important tool. If you can collaborate and coordinate experiments with a like-minded ham at a distant location, you can make some real contributions to the radio art.

8) *A compass and a sextant or a transit*. No, a GPS won't help you here! Doing 3D triangulation of radio sources requires some good old fashioned orienteering or celestial navigation "chops." (It wouldn't hurt to know how to use a compass and a map anyway, in case someone takes out all our GPS satellites).

And a Sidebar

The sidebar on Active Antennas will provide you with some food for thought as well as a great idea for a homebrew project. The active antenna has been greatly undervalued in the ham shack.

The Active Antenna: Real Science in a Small Package

Most radio amateurs and a good number of SWLs are familiar with the so-called *active antenna*. While generally considered a second rate workaround for a full sized antenna, there are aspects of the active antenna that make it a very useful and potentially very precise scientific instrument.

An active antenna is a very short antenna (in terms of wavelength), usually, but not always, in the form of a single-ended whip. (A circularly polarized active antenna is shown in **Figure 9.4**.) Generally less than a few percent of a wavelength at the highest frequency of operation, the active antenna has a radiation pattern approaching that of an isotropic radiator. Directivity is essentially impossible with a single active antenna, but may be achieved with an *array* of active antennas. And, it should go without saying that the active antenna is a receiving antenna only. In the following paragraphs we will compare the active antenna with “normal” receiving antennas.

Figure 9.4 — Roger Weggel, WL7LU, and Clayton Cranor, KL3AB, inspect a prototype “eXOgon” HF CPOL active antenna, designed by the author.

In a conventional full-sized resonant receiving antenna, the goal is to transfer as much power to the receiver as possible. An approximate, if not precise, impedance match is usually desirable, as it is the intent to extract as much power from the ether as possible, and direct it into the front end of the receiver. A perfectly matched receiving antenna will transfer half the intercepted power to the load (the receiver) and reradiate the other half. This is the absolute *best* you can do with a receiving antenna of any kind.

An active antenna works in a very different manner. First of all, consider a “whip” on the order of 1/100 of a wavelength. The radiation resistance will be essentially nothing, while the capacitive reactance will be huge. We could use a very large loading coil to bring the whip to resonance, but in this case it would be extremely narrow-banded. This is essentially the situation with an AM car radio with a normal vehicle antenna. But we generally use active antennas because they’re extremely broadbanded. We make no effort to resonate the antenna *or* match its impedance to the receiver. This would be counterproductive in addition to being an exercise in futility. The active antenna should be thought of as a pure voltage probe, the impedance being so high as to have essentially no current flowing. The power transfer from the antenna is also basically zero.

Now, this brings up a reasonable question, *how does it work at all?*

The trick is the amplifier. If we look at any amplifier that has an infinite input impedance, and an output impedance of anything *less* than infinite, the power gain is *infinite*! If this seems to defy logic, just answer the question, “How much input power can you put into a circuit with infinite input impedance?” The answer is zero. How much power do we get out of such an amplifier? Well, if we have any voltage and any current, we have *some* power. Well, power gain is defined as output power divided by input power, so if we have any output power at all with no input power, we have infinite power gain! This is kind of what we want, to compensate for an antenna with no power to offer, isn’t it?

Now, in reality, the amplifier associated with an active antenna does not have truly infinite input impedance, but it’s very, very high — on the order of several gigohms. A MOSFET or a MOSFET input op-amp will have this kind of input impedance. Typically the first stage of an active amplifier is a source follower FET which has extremely high input impedance and a fairly low output impedance. The output voltage will generally be slightly lower than the input voltage. So we have gobs of current gain, and basically no, or slightly less than unity voltage gain. This still results in tons of power gain…thus sensitivity.

In addition, a source follower has very wide bandwidth, which is something we’re after too.

Now, there are a few caveats when assembling an active amplifier. Since the antenna reactance, and thus its total impedance is so huge, *any* parallel capacitance will essentially short out the antenna. You cannot use normal coax cable between the whip and the electronics. In fact, you shouldn’t use *any*. The whip should be connected directly to the gate of the device, or through an appropriate *series* coupling capacitor. Any unnecessary stray capacitance to the chassis must be avoided. This includes even the feedthrough insulator from the antenna to the amplifier.

Once you have an active antenna constructed and tested, it will be a very valuable measurement tool, in conjunction with your receiver. You can measure *electric field voltage* directly, with essentially no disturbance of the same. Being very close to an isotropic radiator, you don’t have to worry about aiming it, either.

Free Electron Propagation

For the VHF and UHF enthusiast, some of the most interesting propagation is the result of high concentrations of almost nothing but free electrons, or "electron cloud" propagation (**Figure 10.1**). As we've touched on earlier, the "normal" ionosphere is a soup of different species of particles, and even though electrons are the primary ingredient of propagation, we seldom have regions of *nothing but* electrons. Phenomena such as refraction and dispersion are the result of the mixture of ions and free electrons. Things can behave quite differently when we have large electron clouds, and some of the most intriguing…and baffling…beyond-the-horizon communications on the shorter wavelengths are the results of these electron clouds.

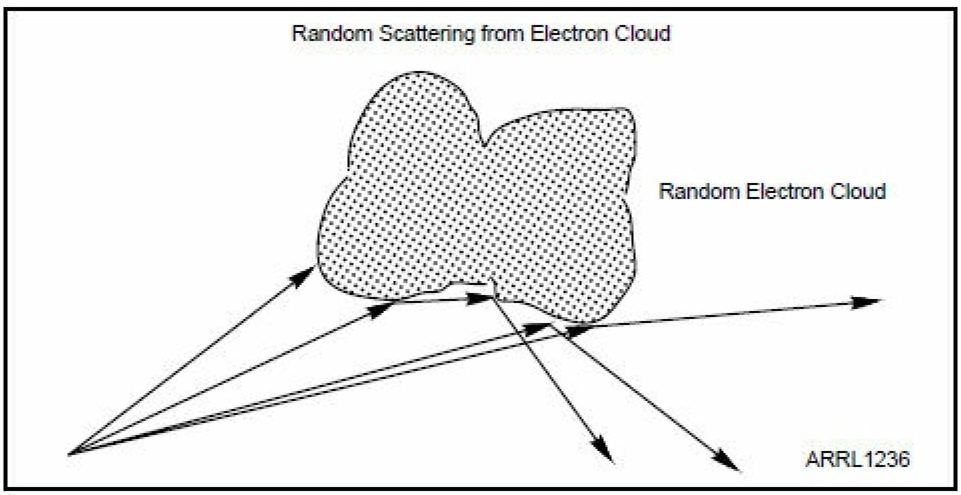


Figure 10.1 — Random scattering off of electrons is an extremely inefficient mode of radio propagation, but nearly always present to some degree. Scattering modes can be observed everywhere from the lowest HF frequencies up to the highest microwave bands. Electron clouds can result from a multitude of sources: micrometeor ablation, local lightning storms, wind shear, and simple random "clumping" of electrons from other processes. Clouds of nothing but electrons, regardless of their sources, are short-lived, because the electrons will repel each other, causing them to dissipate into the neutral atmosphere.

How Free is Free?

We described in our discussion of ionospheric propagation how the electron density profile is related to most factors of HF propagation. In the F region, the electron density profile closely tracks the ion density profile, suggesting that the electrons never stray too far from the atoms from which they were originally torn. It is the presence of the nearby ions that keeps the electrons in something resembling layers, while still allowing them the degree of freedom they need to respond readily to radio waves.

How far can an electron, or a whole mob of them, be removed from their associated ions and still maintain enough collective integrity to work for radio propagation? This is an important question, and one that's coming under a lot of scrutiny by, in particular, meteor burst aficionados.

All electrons *come* from somewhere. To the best of our knowledge, on the grand scale of things, the universe seems to have the same number of electrons as protons. Or, put another way, the overall electric charge of the universe is pretty *neutral*. While most of the universe seems to be in a plasma state, it seems to be a relatively balanced plasma, which is probably a good thing. Large electrical charges seem to be fairly localized.

When the Earth gets a blast of solar wind, likewise, we don't seem to build up huge net electrical charges overall. Otherwise we'd have a really bad static problem on the surface of the Earth — probably sufficient to preclude any life!

The Earth is bombarded by untold trillions of micrometeors every day. Some of these micrometeors are no bigger than talcum powder. To have a *visible* meteor on a dark cloudless night requires nothing bigger than a grain of sand to enter the atmosphere. When these micrometeors "burn up" or *ablate*, a couple of things happen. They can form compounds with the air, and they "rub off" free electrons. As it turns out, the chemistry can be fairly complex, but one of the possible outcomes is that the chemical reaction creates a new *neutral* compound, leaving some free electrons with no *affinity* to anything. In other words, unlike in the F region, there are no ions left to "reclaim" these electrons. So, for all practical purposes, we do have clouds of truly free electrons…but interspersed with neutral atoms or compounds.

Now, we know that a cloud of electrons will rapidly disperse, because electrons "want" to repel one another. We also know that in a plasma, the electrons will only drift so far until they are pulled back home by their "mother" ions. But what about neutrals? Well, as it turns out, a neutral interposed between two electrons has a partial shielding effect. While neither attracting nor repelling the electron, it can also prevent the electrons from repelling each other.

What this means is that you can have a relatively stable, long lasting cloud of electrons that, as far as a passing radio wave is concerned, looks like a clump of truly free electrons. There seems to be some pretty hairy statistical math to describe how well this shielding works…or doesn't. On the whole — again, statistically — it does look like a stable free electron cloud. This is just what we need for a fairly low-loss radio reflection, or rather, absorption and re-radiation.

Another source of free electrons is lightning. We also have some chemistry happening during a lightning storm, but somewhat different from micrometeor ablation. One of the more common reactions of a lightning bolt "burning" air is the generation of ozone, among other molecules, which also leaves some free electrons. However, lightning is an interesting situation, because in order for there to be enough charge to create a strike in the first place, there has to be some aggregation (pre-ionization) of the molecules (most frequently water) to start with. So, it looks like lighting itself is merely a second source of ionization; most likely it is *wind shear* and similar atmospheric effects that create the energy necessary for ionization in the first place. The jury is still out on this.

What We Know

One of the certainties of "free electron" propagation is that is always occurs at relatively low altitudes, such as the D and lower E region. It is only at these altitudes that we have enough neutrals to perform the "shielding" function, as well as to supply electrons in the first place, without radiative input.

Because E layer propagation is frequency insensitive, it is most amenable to VHF propagation where high gain, narrow beam antennas are practical. By using careful triangulation methods (in three dimensions), we should be able to determine fairly accurately the locations of electron clouds. This is an area few hams are exploring. One of the big mysteries of sporadic E is just how big (or small) the clouds are that allow such effective communications.

This is also an area where we could be using a lot more of our microwave bands. The upper troposphere and lower ionosphere are regions that have strong effects on microwave frequencies. And it is quite likely we could see some frequency dependent effects in the upper microwave region, but again, this is pretty sparsely populated, as far as hams are concerned.

Auroral Matters

Auroral phenomena may also be partially the result of free electron masses. Auroral propagation can be in effect throughout the upper HF, VHF, and UHF frequency ranges. The visible aurora actually *begins* at E level altitudes and extends well into the F2 region. People are quite surprised when they learn that a "full curtain" aurora can be several *hundreds of miles* tall! Because the auroral region covers a vast region of sky, it is difficult to categorize auroral propagation into just one mode. However, we can make a few generalizations. Visible auroras are the result of charged particles, and since these particles can be infused with great intensity and suddenness, we can expect large congregations of free electrons to be present, if short-lived.

These electrons, if there is a strong magnetic field nearby, will become part of *plasma tube* structures which have extremely well-defined borders. Imagine a sheath of free electrons enclosing a thin core, somewhat like a coaxial cable. So called *field aligned irregularities* (FAIs) are very easy to see as you sweep a VHF or UHF signal across the region. UHF *radar* methods show these FAIs even more clearly. So what we have in this case is actually a free electron structure embedded within a more normal F-region plasma soup.

It should be noted here that auroras can occur any time day or night, at any time of the year. They are clearly visible on HF radar in interior Alaska, even during the long summer of the midnight Sun. **Figure 10.2** shows some vintage returns from a 28 MHz auroral radar. And, as mentioned above, the auroral region, being as tall as it is, forms a fairly good sized billboard reflector even for upper HF frequencies. Auroral propagation may account for a lot more HF than we commonly give it credit for.

More Oddities

As we've suggested, fast-moving charged particles are *in general*, counterproductive to good radio. As with any great general principle, it's the exceptions that make it interesting!

To reiterate, fast-moving charged particles disturb the geomagnetic field, which "wrinkles" the ionosphere, thus spoiling its smooth reflective and refractive properties, among other things.

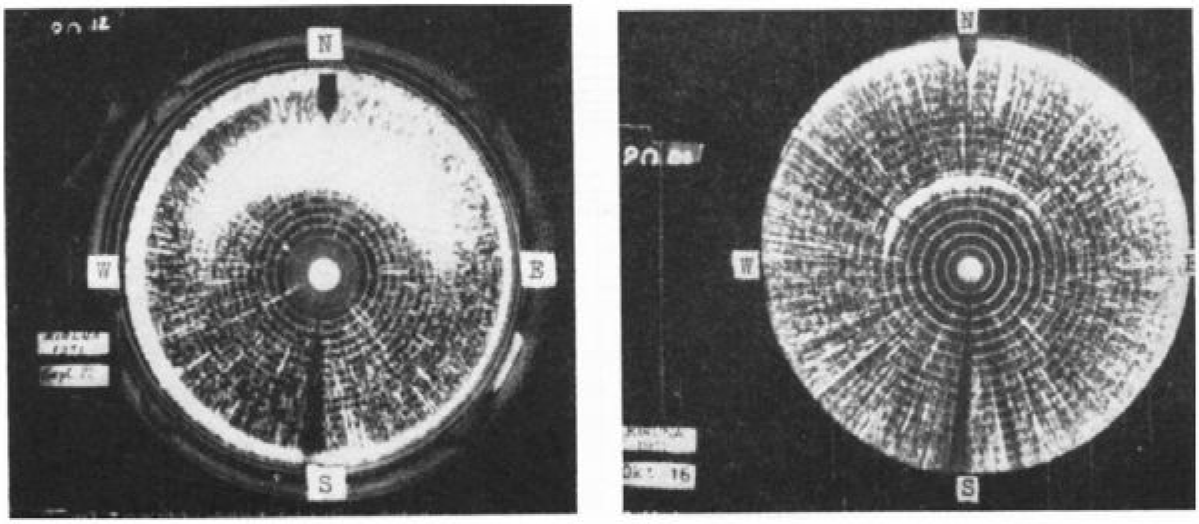


Figure 10.2 — Back in the 1950s, 10 meters was commonly used for auroral radar. The image on the left shows radar returns from an auroral curtain and the image on the right shows radar returns from an auroral streamer. [courtesy Richard Dickman, former site manager of Hipas Observatory]

On the other hand, clumps of slow moving, or even stationary electrons give us all kinds of interesting surfaces off of which to bounce radio waves.

The good news in all of this is that fast-moving charged particles eventually become slow-moving charged particles. This is why you sometimes have some wonderful — and peculiar — propagation events immediately following a near complete radio blackout as one might encounter after a coronal mass ejection (CME), where the Sun "blows chunks."

Remember how we described just how sparse and tenuous the ionosphere really is? The fact that we have layers at all is actually a pretty delicate process, and it takes time for things to settle out after a disruption to the point where we have any recognizable structure. A major solar disruption can totally stir up the ionosphere, blending all the layers into an amorphous colloidal suspension-like blob. Fortunately (not just for radio, but probably for life itself), the ionosphere (and all our other spheres) are fairly self-correcting. You can take water, salad oil, honey, and gasoline and put them all in a blender, and come out with a grayish-yellow milkshake or something. (We don't recommend drinking this.) But after a while, you will have some layers forming, because of the different densities of the liquids. With a little stretching of the imagination, the ionosphere is a bit like this.

Now, in the intermediate stages between total chaos and normal layering, we can have some interesting effects. One phenomenon is known as *electron precipitation*, which is exactly what it sounds like (see **http://blogs.agu.org/geospace/2010/12/14/its-raining-electrons/**).

After a major auroral event, a lot of free electrons fall out of the "milkshake" and follow the magnetic field lines to the surface of the Earth. (In places like Fairbanks, Alaska, you can actually measure this electron precipitation with a simple old-school *electroscope,* after a strong visible aurora.) Since these precipitating electrons are fairly slow (they *can* be, at least, although there are also energetic precipitating electrons), they can be useful for bouncing radio signals.

Rain Man

Water is capable of taking on a tremendous electrostatic charge. Water droplets accumulate charges on their surfaces, primarily because water is a pretty bad conductor. If you have a lot of finely atomized water, you have a lot of surface area upon which to accumulate a static charge, i.e., electrons. A charged water droplet is basically a bag of free electrons for the taking, except that the bag is on the inside. What this suggests is that we should be able to use clouds as highly reflective radio surfaces…if we choose the right frequencies. You don't want to pick a frequency near the resonant frequency of water, so you want to avoid the "microwave oven band" near 2.45 GHz. Fortunately, the closest ham bands we have to that are 2.300 – 2.310 GHz and 2.390 – 2.450 GHz. All our other microwave bands are far enough away to not be too absorbed by resonance effects. This leaves at least 10 other bands with good potential for rain cloud bounce. Again, we really need to activate these "orphan" bands to explore this sort of thing.

Of course, not every cloud is highly charged. Neutral clouds floating around probably would not be of much use in this application. But what this does potentially give us is the ability to predict lightning storms by pinging suspect clouds with various microwave frequencies and seeing how they respond; a lightning cloud about to "blow" should be highly reflective.

Incidentally, this is a bit off topic, but maybe of some interest to meteorologically-minded hams. For many decades, it was assumed that rain clouds always needed some kind of "seed" for raindrops to form. The semi-ancient process of "seeding" clouds with iodine crystals and other chemicals for the purpose of making it rain was never much questioned. However, recent research in the plasma lab fairly conclusively shows that the *main* thing keeping water vapor aloft in cloud form is *static electricity*. If you discharge the vapor, the surface tension is broken and the vapor can now condense into larger droplets. (You can make absolutely pure distilled water "rain" inside a plasma chamber by discharging an electrically suspended cloud, thus obviating the need for any kind of seed.) This also largely explains why it so often rains directly after a lightning strike. Now you know.

A Bit Higher, a Bit Weirder

While cloud-bounce is most likely the realm of UHF and microwave, there are a couple of mysterious atmospheric phenomena that are highly reflective of HF and even medium wave frequencies. These are known as *noctilucent clouds*, and *polar mesospheric summer echoes* (*PSMEs*) which many scientists believe are actually the same thing, just at different seasons. This is still under intense investigation.

PMSE's are below the D layer, in the no-man's land between the stratosphere and the ionosphere. When they are in "action" they can appear as fully *specular* reflections at frequencies as low as 2.5 MHz, and possibly lower. (At Hipas Observatory, we would convert our HF heater array to serve as a medium wave vertical incident radar to investigate these). They can come and go within seconds, but when they exist they are unavoidable, creating near total reflection of vertical incident radio signals. Needless to say, in order to provide such an effective reflection for medium wave signals, they have to occupy a *lot* of sky. And, presumably, they have to have extremely high concentrations of free electrons. As of this point, the only active radio investigation of these involves *vertical incident* probing. Oblique (low angle) sounding of these has not been performed in the scientific community, and may be something of interest for Amateur Radio investigation. Imagine *really* short skip on 160 meters! Although these were once considered a purely arctic phenomenon, there is some evidence that whatever they are, they're gradually expanding southward as well.

How About Free Proton Propagation?

Since it's a given that all electrons have to come from somewhere, it goes without saying that for every free electron we have, there must be a free ion someplace. Can we likewise use ion clouds to bounce radio signals from?

Theoretically, yes; practically, no.

As we know, every proton has the same charge as every electron. But the proton has 2000 times the mass…and to add insult to injury, the ion clouds that we have in the ionosphere and elsewhere are not likely to be proton clouds, but clouds of much heavier ions.

Recall our discussion earlier on *nonlinear* ionospheric phenomena. Now, while we may not be able to wobble an ion enough to get a meaningful level of RF re-radiation, it *is* possible that we might be able to instigate something known as *ion acoustic waves.* Ion acoustic waves are essentially sound waves, but rather than being the result of collisions between neutrals, they are the result of propagated repulsive nudges. As such, they are *compression* (longitudinal) waves, and they travel much more slowly than electromagnetic waves, although much faster than sound waves at ground level. It is conceivable that ion acoustic waves are a component of long-delayed echoes and similar phenomena. The jury is still out on this.

Try it and See

As with every other phenomenon we encounter, there is always room for experimentation. We still have a lot to learn, and as a radio amateur you are in the perfect position to do just that.

Neutral Propagation

By *neutral propagation,* we mean propagation through normal *air.* For the most part, Amateur Radio literature (and commercial radio literature, for that matter), has dealt with propagation through air as being essentially the same as propagation through a vacuum. However, there are enough *observable* effects of radio propagation through "normal" atmosphere to merit giving the matter some attention.

Dielectrics and Other Delights

The dielectric constant of air at one atmosphere of pressure is 1.00059. If a pure vacuum has a dielectric constant of 1.00000, it's pretty clear that air should act pretty much like a vacuum, as far as radio is concerned. But let's take a look at the actual process by which a dielectric affects radio propagation. While the terms *insulator* and *dielectric* are often used interchangeably, they are actually two different properties. While a good insulator generally has a high *dielectric constant*, these can both exist independently. An insulator is a material with low electrical conductivity (or inversely, very high resistance). A dielectric, while almost always being a highly resistive material, also has a *relative permeability* which is the ability of the atoms to be *polarized.* A vacuum has high resistivity because there is nothing in it to conduct electricity, and it has a low dielectric constant because there's nothing in it to polarize!

Air, likewise, is a good insulator, because there are no free electrons (unless it's been ionized), and at atmospheric pressure there aren't many molecules around to be polarized. So it also has a very low dielectric constant (though slightly more than a vacuum).

When we send a radio signal through a volume of air, the atoms will align themselves with the electric field of the electromagnetic wave…that is, they will shift their position slightly within the bulk material. But, you may protest, "How can the molecules respond at all to a radio wave since they're neutral?" Didn't we long ago establish that only charged particles respond to radio waves?

This is absolutely true, when we look at the atom or molecule from a distance (**Figure 11.1**). But any atom has *asymmetries* at some point in time. If we take a single hydrogen atom, for instance, with one electron circulating one proton, we still have this asymmetry. An electric field applied to the hydrogen atom will apply an equal and opposite force to the proton and the electron (**Figure 11.2**) — but not at the same time! This is because the electron is already moving around the proton. So, the field will still apply a "bias" to the entire system.

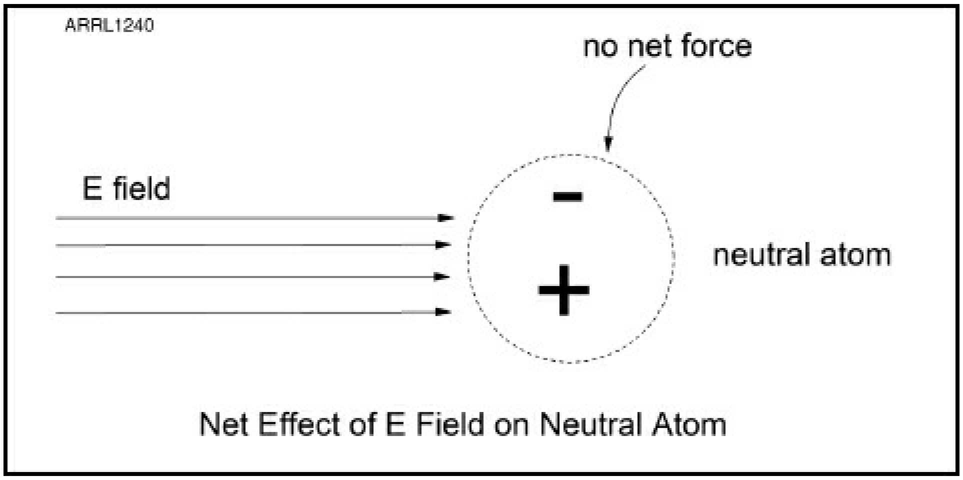


Figure 11.1 — A neutral atom will generally be invisible to distant electromagnetic fields because the forces upon the opposite charges are equal and opposite.

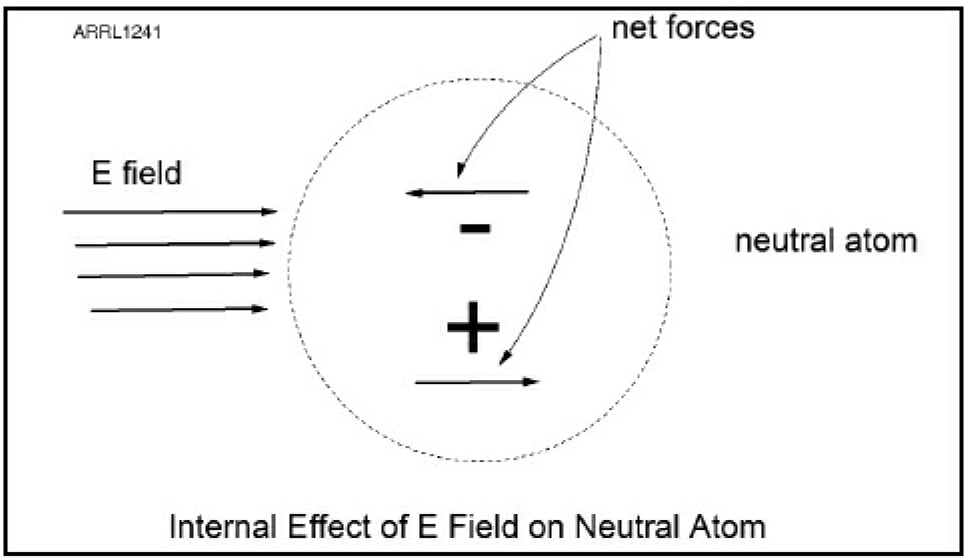


Figure 11.2 — Although to the outside world there is no net acceleration of a neutral from a radio wave, there is an effect *internally.* Radio waves can cause neutrals to spin, for instance.

Now, this is a bit oversimplified, since in air we have not only molecules, but different types of them. Nevertheless, the electric field passing through the midst of a "chunk" of air will offset the molecules, either in position, or in rotation…or sometimes both. This process is not instantaneous. Since the particles have mass (or angular momentum) it takes a finite time for them to shift to their new polarized state. When the next half cycle of RF passes through, they not only have to return to their original state, but pass on through to the *opposite* polarized state.

But that's not all! Remember the universal concept of reciprocity? Every accelerating charge creates an electromagnetic wave and every electromagnetic wave will move an electric charge. This applies here as well. As the atom or molecule returns to its non-polarized position, it creates a small electromagnetic wave, which contributes to the original wave. The sum of the original wave and the "re-radiated" delayed wave contributed by the polarized molecules creates a new wave which is slightly slower than the original wave.

This process can be either lossy or lossless. In other words, the polarization can be entirely *elastic,* wherein all the energy is returned to the original wave, but slightly delayed in time. Or the process can be inelastic, where the polarization results in heat loss (friction). The amount of loss is rather complex, and depends on temperature, chemical composition, and other factors. In a dielectric, this heat loss is sometimes referred to as a *loss tangent*, and mathematically shows up as an electrical resistance. You're probably asking yourself about now, "How can the loss tangent increase the resistance, when it's already in a perfect insulator?" This is an excellent question. The answer is that there are countless *parallel* resistances in a real dielectric material, and a lossy resistor in parallel with an open circuit is still a lossy resistor! (This also explains why you can sometimes *increase* ground losses of an antenna by adding *more* radials! An open circuit has less heat loss than a lousy circuit!) We'll discuss this a bit further in Chapter 12, Cheaper Than Dirt.

Mirages and Such

As we have seen, things are not always as they seem. A "perfect" insulator can have an effect on radio waves. Although we have countless different dielectric materials, air is somewhat special because the vast majority of our Amateur Radio communications take place in air. We know that air has varying densities as well, dependent on altitude, base barometric pressure, and temperature. Water content has an effect, as well.

The dielectric constant of air changes with respect to pressure, as well. Under typical conditions, this change of pressure is so gradual that we don't see much effect on radio signals of any frequency. There will be a small amount of dispersion of a VHF signal as we move from ground level to the stratosphere, but barring any unusual activity, such as sporadic E, until we get to the ionosphere, radio propagation is nearly the same through air as in a perfect vacuum.

However, *sudden* changes in pressure (or density) can result in sudden changes in dielectric constant and refractive index. When a radio signal encounters an abrupt border between air of one refractive index and air of another, we can have substantial refraction…or even reflection… occurring. We've all seen *mirages* over a road on a hot day, caused by the abrupt change in density of air directly over hot pavement to that of air a few feet higher. Objects far beyond the normal horizon can be seen clearly. Radio is electromagnetic, just like light, and can undergo the same effects. Radio "mirages" can occur at VHF, but even more dramatically at microwave frequencies. These reflections can actually become quite problematic for commercial microwave links. If they happen to arrive out of phase at a receiving antenna with the direct wave, total blackouts can occur.

"Clear air" reflections can also occur in the case of a strong *inversion* layer where a region of high pressure sits on a lower pressure region. In optics, we have "superior" and "inferior" mirages, which are reflections from an inversion "ceiling" and an inversion "floor" respectively. Likewise we can have either type of radio mirage or sometimes both simultaneously.

The Art of Being There

Predicting conditions amenable to over-the-horizon tropospheric propagation is very… well….*unpredictable.* Since these events are so closely tied to weather phenomena, their prediction is equally unreliable. It has been noted that though scientists have been predicting weather for millennia, while ionospheric physicists have less than a century under their collective prognostication belts, ionospheric prediction is generally much more accurate than weather prediction. To their credit, meteorologists don't have the advantage of advance warning that ionospheric physicists have. Solar conditions have somewhat direct and known effects on the ionosphere, while there is no such singular predictor for local weather.

The take-away from this is that hams who are successful at taking advantage of the myriad of exciting tropospheric events are successful because they are around when they happen. Such events can be rare and short-lived. Even if there were more reliable methods of predicting them, by the time you get the word, the event could be here and gone. This is not only true of tropospheric events, but also sporadic E, which again is called sporadic E for a reason!

Complicated

By now, it should be somewhat apparent that no mode of long distance radio propagation is mutually exclusive. There is nothing to prevent a radio signal, say, from undergoing sporadic E for one leg of a journey, followed by a nice F1 hop, finishing up with a conventional F2 hop. Multimode propagation paths can exist from HF frequencies clear up to the microwave bands. It is this fact that accounts for a lot of Amateur Radio discussion, when it comes to identifying just what path a radio signal is taking from Point A to Point B. This is also a great part of the intrigue. By the way, here is a neat overview of some "multimode" propagation paths, by the National Institute of Standards and Technology (NIST), formerly known as the National Bureau of Standards: **http://nvlpubs.nist.gov/nistpubs/sp958-lide/120-122.pdf**.

It is almost certain that so called transequatorial propagation is just such a multimode phenomenon. We'll discuss transequatorial "skip" in more detail later in a more appropriate chapter, as it is definitely *not* a neutral type of propagation.

All Wet

No discussion of neutral propagation would be complete without talking a bit about fresh water. We specifically mention *fresh* water, because salt water (which is most of the Earth's surface) has a profoundly different effect on radio than fresh water.

As with any dielectric medium, fresh water has two fundamental electrical properties, namely electrical resistance and dielectric constant. Curiously enough, the dielectric constant of salt water is essentially the same as that of fresh water, while the conductivity is very different. Sea water is highly conductive, while distilled water is nearly a perfect insulator. But not quite.

One thing that makes water so unique (and it's somewhat paradoxical that a substance that is so prevalent on our planet is also so *weird)*, is that it is a polar molecule. Water has a property known as self-ionization, which is a relatively weak property and not of great interest as far as radio propagation goes. However, it has a very high dielectric constant at room temperature of 80.18. This means radio waves slow down a *lot* when they pass through a volume of liquid water. However since pure water is a very good insulator, the losses are generally negligible. HF and VHF frequencies can easily penetrate long distances through fresh water. A couple of scuba divers could easily use a pair of 2 meter transceivers to communicate with each other underwater — if they were in a fresh water fish tank. However, it would be nearly impossible for a person above the tank to communicate to a scuba diver in the tank, because the abrupt change of dielectric constant from air to water will cause nearly total reflection from the surface. Any signal that *did* manage to penetrate the surface would be sharply bent *below* the surface.

The bottom line of this is that large bodies of fresh water, such as the Great Lakes behave pretty much the same for reflecting incoming skywave signals as do the oceans. However, for low angle *ground waves*, the Great Lakes and the oceans are very different! We'll talk more about the "ground wave" properties of water in the next chapter.

Cheaper Than Dirt

Most professional and amateur antenna literature discusses the role of the immediate ground conditions on the performance of an antenna, but little is generally said about the long distance effects of the ground on radio propagation. In actuality, the ground has a great deal to do with both the antenna characteristics and the distant propagation characteristics. More recent antenna modeling programs do a much better job of describing the "far ground" characteristics than those of the recent past, so we'll mention a few words about modeling the ground. But first, let us look at some basics.

Look Out Below

Unless you are working with upper UHF or microwave frequencies, most Amateur Radio antennas are going to be close enough to the ground, relative to wavelength, that the effects of the ground are non-negligible. In a majority of cases, the term "non-negligible" is a gross understatement; most hams severely underestimate the effect of the ground underneath their antennas, regardless of whether they're working MF, HF, or even VHF. Of course, the effects of the Earth are progressively less as you increase in frequency, for a given antenna height, but there can occasionally be surprising ground effects even at microwave frequencies under the right (or wrong) conditions.

In keeping with good scientific methodology, it's best to explore the *normal* before attempting to explain the extraordinary. As a first order approximation, the ground underneath an antenna can be thought of as an infinitely large plane reflector. The quality of this reflector is highly variable, but it can be helpful to look at it as being a perfect mirror to start with. As a benchmark, let's consider a 40 meter horizontal dipole, erected at a height of 1⁄4 wavelength, or 10 meters over the ground. This was a pretty common "default" antenna for a few generations of impoverished radio amateurs wanting to get on the air as quickly as possible. Most hams could get a wire antenna 30 feet or so into the air, suspended by a couple of appropriately spaced trees, typically by bribing a younger sibling to risk life and limb (both of the tree and his own) to climb up and tie off the ends of the antenna to a couple of accommodating branches.

This is actually a rather effective arrangement, for a number of reasons. The effective radiation resistance of a dipole over real ground is the product of its free space radiation resistance (theoretically 72 Ω) and the contribution of the ground reflected signal. This ground reflected signal is dependent on both the reflectivity of the ground and the effective height. Interestingly enough, at 1⁄4 wavelength, the radiation resistance for both the *perfect ground* state and the *real world* ground state is nearly identical, and quite close to 72 Ω. This is shown clearly in **Figure 12.1.** As a result, this antenna presents a good impedance match to every modern transmitter with any length of coaxial cable.

In addition, the horizontal radiation pattern, resulting from both the direct radiation and the ground reflection is a good compromise, giving a fair balance of low-angle radiation and high angle radiation. Some skeptics describe this general antenna as working "equally poorly in most directions, and not at all in others," but it deserves more credit than that, actually. One surprising effect of the ground underneath the dipole is that it now has radiation off the ends, but this radiation is *vertically polarized*! This is clearly shown in any *NEC* antenna modeling program, but also borne out in experience. Very few hams report perfect nulls (or even any very deep nulls) from any direction (in azimuth) when using a simple HF dipole, which would certainly exist if in free space! (To be fair to the skeptics, this antenna *does* indeed show a deep null at very low *elevation* angles, nearly regardless of azimuth). But the overall analysis is that this is a pretty good antenna.

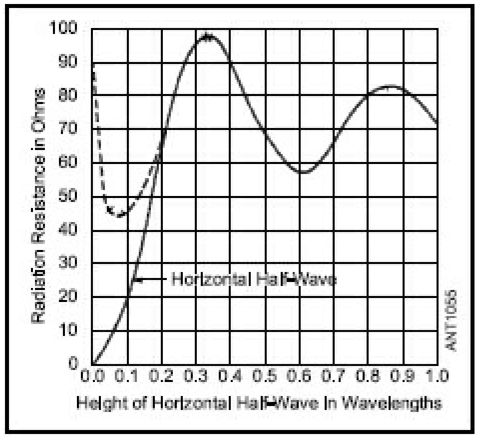


Figure 12.1 — Curves showing the radiation resistance of a horizontal half-wavelength dipole at various heights above ground. The broken- line portion of the curve for a horizontal dipole shows the resistance over *average* real earth, the solid line for perfectly conducting ground.(courtesy the *ARRL Antenna Book*, 22nd Edition, Chapter 2)

Contributions, Please

At this time we would like to again plagiarize some passages from the *ARRL Antenna Book*, but add some further relevant commentary. These passages address the contribution of the Earth to radio propagation, especially to horizontally polarized waves. Again, from the 16th Edition, Chapter 3:

"In considering earth characteristics, questions about depth of RF penetration often arise. For instance, if a given location consists of a 6-foot layer of soil overlaying a highly resistive rock strata, which material dominates? The answer depends on the frequency, the soil and rock dielectric constants, and their respective conductivities. The following equation can be used to calculate the current density at any depth:

Knowing σ and e_r, the skin depth in an arbitrary material can be determined from:

$$\delta = \left(\frac{\sqrt{2}}{\omega\sqrt{\mu\varepsilon}}\right)\left[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2}-1\right]^{-1/2}$$

where

δ = skin or penetration depth [m]

ω = 2pf, f = frequency [Hertz]

σ = conductivity [Siemens/meter, S/m]

$\mu = \mu_r\ \mu_o$ = permeability

μ_o = permeability of vacuum = $4p10^{-7}$ [Henry/meter]

m_r = relative permeability [dimensionless]

$\varepsilon = \varepsilon_r \varepsilon_o$ = permittivity [Farad/meter]

ε_o = permittivity of vacuum = 8.854×10^{-12} [Farad/meter]

ε_r = relative permittivity [dimensionless]

By some manipulation of this equation, it can be used to calculate the depth to which the current density is some fraction of that at the surface. The depth at which the current density is 37% ($1/e_0$ of that at the surface, often referred to as *skin depth*) is the depth at which the current density would be zero if it was distributed uniformly instead of exponentially.…"

Notice the beloved term "skin depth" which pokes up its pretty little head whenever we deal with radio frequency electric currents. While the moderately horrendous equation presented above is, well, moderately horrendous, the simplified implication is pretty straightforward. Electric currents like to travel on the outside of conductors, and whether that conductor is a piece of wire or a piece of real estate, the principle is the same. The higher the frequency is, the less the current is going to penetrate the conductor. Lower radio frequencies penetrate further into the Earth (or the ocean) than higher frequencies.

Now, we know that the ultimate radio frequency signal strength at a distant receiver is the result of the contribution of all radiators, and is a function of both the amplitudes of all contributing sources of radiation and their relative *phase angles.* A contributing source can be a radiator, a reflector, a director, or even a "virtual element" such as one might see from a ground reflection.

The Skinny on Skin Depth

As one might suspect, a piece of copper wire, consisting of a uniform mass of a single element, is a lot simpler to analyze than *dirt*, especially in regard to skin depth. Unlike a piece of wire, the actual location of a reflecting "wire" under the ground is not easy to define. An in the passage above, we see a continuous *gradient* of currents when a radio signal penetrates the Earth (**Figure 12.2**).

What we really want to know is where the greatest *current density* is, because a region of greatest current density most closely approximates a wire. And, if we have the patience to actually apply the aforementioned moderately horrendous equation to all our known dirt components, we could accurately pinpoint this location. We could find an actual depth of our reflecting underground virtual wire. What we would end up with is a *virtual depth* of our ground. And, as it turns out, this depth can be…well…very deep. And, in fact, in *most* locations, this virtual depth is a lot deeper than most hams realize. This is why a dipole antenna electrically appears to be, in most cases, higher than it actually is. Or, put another way, low dipoles generally work "better than they should." This is good news!

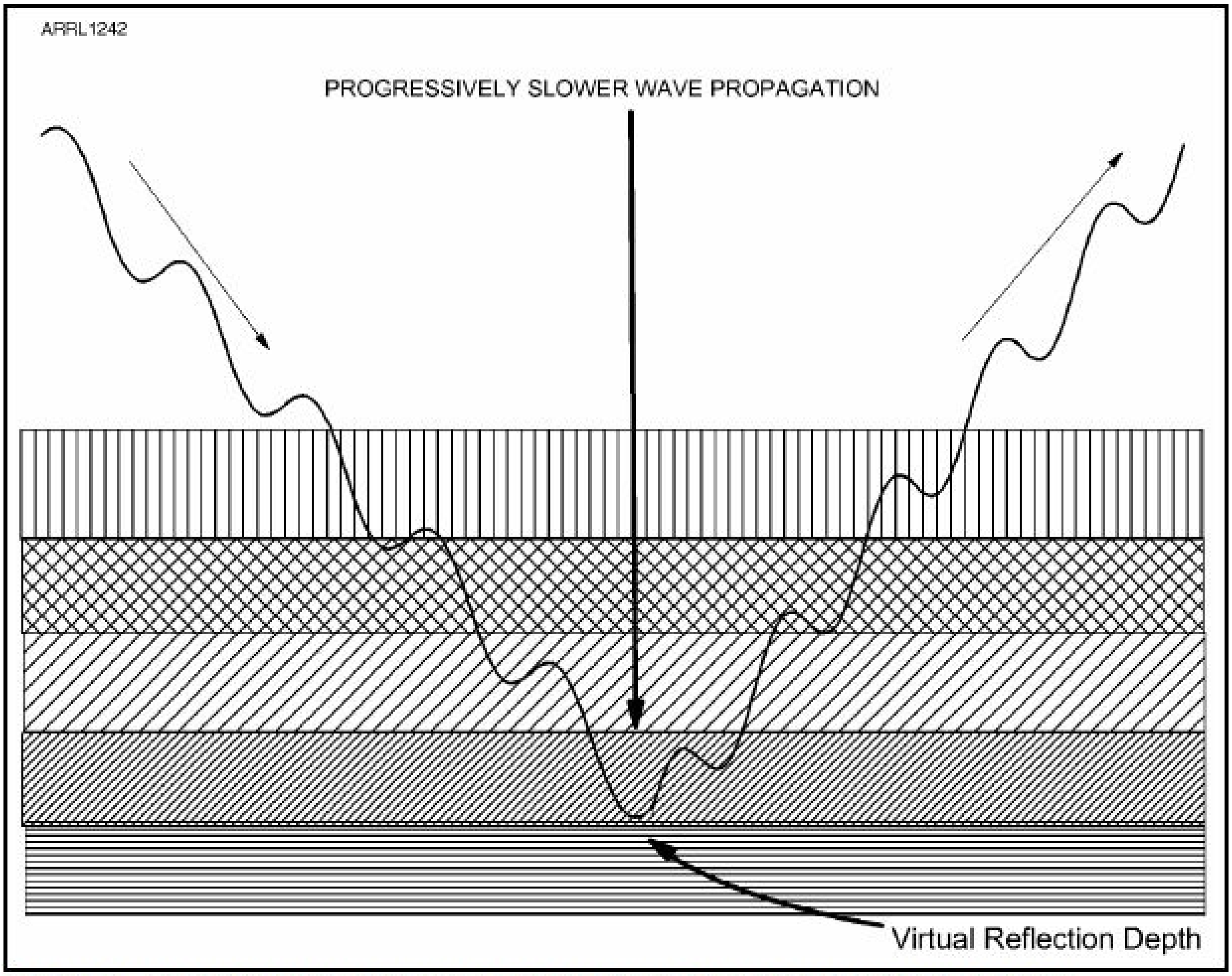


Figure 12.2 — Real earth has some similarities to the ionosphere. Dirt can have many layers of different conductivities and dielectric constants. In most cases, at least at HF frequencies, the depth of penetration can be significant. In addition, because of the sometimes high dielectric constant exhibited by soil, the radio waves slow down significantly, making the *virtual* reflection depth much deeper.

Now, if you've been following carefully, you might have discovered a bit of a paradox with this skin effect business. We know that, at least for a uniform medium, the current density decreases with depth. So, intuitively we think that the further down you go, the greater the losses are. But remember losses are subject to Ohm's Law — the "I Squared R" law rules supreme. For a given value of resistance, the greater the current, the greater the losses. So, based on this, it would seem that you'd have the greatest loss where the greatest conductivity is, because that's where the greatest current is…and we have already established that this is at the surface of a conductor. So does this mean that a lousy conductor has less loss than a good conductor at radio frequencies?

As it turns out, there is a balancing point. For those of you who have worked with high powered radio transmitters for a while, you know that it's standard practice to silver plate RF coils and cavities, because you want the highest conductivity on the surface where the greater currents are. This follows our reasoning with regard to "I squared R" losses. But skin effect is a two-way street. We can *affect* where the maximum current density exists by changing the conductivity of the material!

And this is the primary "hitch" when dealing with a real Earth consisting of different layers of differing conductivity. The differing layers can "massage" the current density contour to a region that seems to defy the normal skin effect calculations.

We also need to rethink the common "wisdom" that electricity always "takes the path of least resistance." The truth is that electricity takes *every possible path*. The only truth to the aforementioned "wisdom" is that the *greatest current* takes the path of least resistance.

Now, let us return to our plagiarism of the *ARRL Antenna Book* and gain some more insights.

Continuing on:

"…The depth in fresh water is about 156 feet and is nearly independent of frequency in the amateur bands below 30 MHz. In salt water, the depth is about seven inches at 1.8 MHz and decreases rather steadily to about two inches at 30 MHz. Dissolved minerals in moist earth increase its conductivity…"

We should note here that the skin depth is not a brick wall. As with any logarithmic function, such as the familiar RC time constant, or the inverse square law of radiative power, it gives us a useful line of demarcation between "good" and "bad." And 1/*e* is a great "cutoff" point, since it exists everywhere in nature. You'd be hard pressed to find *any* physical function that does not follow the 1/*e*. It's hardwired into the universe.

Counterpoint

You may have noticed some striking similarities between the ionosphere and dirt. They are refractive, reflective, and conductive. They are both dispersive…that is, frequency dependent. Recall that frequency is a factor of skin depth, just as it is with the depth of penetration of the ionosphere. Both the ionosphere and dirt are characterized by a complex number, representing conductance and dielectric constant. Of course, the dirt doesn't generally move around as much as the ionosphere, but it does have periodic and non-periodic changes.

Up till now, we've dealt with ground conductivity (or lack thereof) as a necessary evil. We know that things would be much simpler if we did everything in free space. We have learned that we can't simply ignore the ground and its effect on our radio enjoyment. Now we will look on the other side of this, and talk about how to really use the ground to our advantage, not just merely tolerate it.

Multihop

Most of us understand that multihop ionospheric propagation would be impossible without intermediate ground reflections. The one exception to this is the relatively rare *chordal hop* propagation discussed in an earlier chapter.

While most HF propagation prediction tools are good at taking into account the local ionospheric conditions, even through multiple time zones, they don't take into account local or regional variations in ground conductivity, but model the ground reflections on a perfectly conducting sphere. This works reasonably well, but probably is far from ideal. Perhaps what the world needs is a good "multihopagation" prediction program!

What's for Launch?

It may seem a bit illogical to write an entire book to explain phenomena and behaviors of nature of which you *can do absolutely nothing about*! Inasmuch as we're 12 chapters into this exercise, it's a bit late to be thinking about such rhetorical existential matters. We're here now.

Actually, the situation is nowhere near as grim as you may think. As the saying goes, "You can't stop it from raining, but you can bring an umbrella." And, as unpredictable and uncontrollable as radio propagation can be, we do have a lot of umbrellas up our sleeves, and a few other tools we haven't even thought of. Knowing the difference between what you can control and what you can't is often the best tool you can have.

This brings us to the greatly underrated idea of *vertical launch angle control.*

Most radio amateurs, even seasoned veterans, give little or no consideration to the vertical launch angle of their signals. Probably even less attention is given to the *incoming vertical arrival angle* of incoming signals. Let's deal with the former first.

We might as well put it down in writing: Half of any antenna you use is the ground.

This may not be strictly accurate from a numeric standpoint, but since the ground has been *undervalued* for so long, we're probably in no danger of giving it *too much* value, at least in one chapter.

Many well-heeled (or heavily indentured) radio amateurs spend a great deal of consideration and money on optimizing the azimuth gain of their antennas. Naturally, one wishes to aim as much energy as possible in the general direction one wishes to communicate. And, since most hams with which one wishes to communicate are somewhere near ground level, they assume that that is all that matters. A corollary to this is that higher is always better. The composite result of this thinking is that the best antenna is, almost by definition, the most expensive. Getting big heavy objects high in the sky is always an expensive proposition. So is keeping them there.

If you were to ask the average ham why he or she wants to have the antenna as high as possible, he or she will answer something like, "That's a really stupid question. The higher the antenna, the farther your horizon will be!"

Oh really?

A little thought and some simple calculations will reveal just how *quickly* you reach the point of diminishing returns when it comes to horizon distance versus antenna height…*especially* at HF. Remember, the horizon distance only goes up as the square root of antenna height! For all practical purposes, you will gain *absolutely no discernible* horizon advantage over about 45 feet!

"But…but…but…look at all those big gun stations with their 200 foot towers! They always clean up in the contests, so they must be doing something right!"

Certainly, they're doing something right, but not for the reasons you think. If horizon distance were the main issue, then someone with a 30 foot tower in Denver, Colorado, would always clean the clock of any ham in Death Valley, California, even with a 200 foot tower, what with his mile-plus height advantage. We *know* that's not the case!

The *real* advantage of having a really tall tower (at least on the high bands) is that you are up high enough to take advantage of *far field* ground reflections, *plus* being up high enough to avoid lossy ground coupling (near field induction and capacitive losses). As it turns out, you really need to be close to the "legal limit" of 200 feet to take advantage of this. (Amateurs who want to construct an antenna structure more than 200 feet high must notify the FAA and register the tower with the FCC to avoid unknowingly creating hazards to aircraft. Additional restrictions apply if the antenna is within about 4 miles of a public use airport or heliport.)

For simply taking advantage of near field reflections for lowest launch angles, your best height is at about 1⁄2 wavelength, at the frequency of operation. This is obviously a lot lower than often thought. At 20 meters, 1⁄2 wavelength is only about 33 feet…and this is assuming a perfect ground where the reflection is on the surface. But as we've just discovered, the virtual reflection height can be far below "physical" ground level.

There's also another "sweet spot" at about one wavelength of height; you'll have some additional forward gain from the ground reflection, but also a few more vertical "side lobes" which may or may not be helpful.

We've made the assumption here that low launch angles are the desired goal, and for DX chasing and most contesting, this is a fairly good assumption, because, all things being equal, the more distance you get per hop, the lower the path losses. However, you can also end up hopping right over someone, so there are times you want to have a higher launch angle. But how do you accomplish this with a typical Yagi or quad beam?

Well, you *could* put an elevation rotator on your antenna, but how many of those have you seen on HF recently? The most obvious method is to change the height of your antenna, and the crank up tower is ideal for this. Unfortunately, most crank up towers are operated in only two modes — fully up for operation, and fully down for maintenance. Countless hams could gain some real operational advantage by using their crank-up towers as a vertical steering device. Of course, you need to take safety into consideration, as not all crank-up towers are safe in a "half-cranked" situation. And of course contesters have those 200 foot towers so that they can have multiple antennas at different heights.

Diversity Methods

The ionosphere can be fickle. The fact that a large number of hams practically make a career out of attempting to predict propagation (with varying degrees of success) attests to this. The variability of ionospheric (and other types) of propagation has been observed for nearly as long as radio itself. We say *nearly* as long, because for a decade or so after its inception, nearly all Amateur Radio was conducted via the annoyingly consistent ground wave (or “sea wave”).

Part of good engineering practice is knowing what you can fix, and knowing what you can’t fix. Applied to radio propagation, it’s safe to say that there’s really little point in attempting nighttime international communications on 10 meters during a sunspot minimum. It “just ain’t gonna happen.” If there’s no ionosphere, there’s no skip. This is not to discount anomalous events like meteor burst that can happen just about any time. But we can reasonably well predict those times when it’s just not worth the bother to turn on the rig on a certain band.

Diurnal (daily) variations in maximum usable frequency (MUF) follow a nearly clockwork pattern of predictability, as well. We can plan our operating habits around them, such as using known “daytime bands” during the day and “nighttime bands” at night.

On the other hand, there are short-term variations in propagation that can be dramatic…and unpredictable. We refer here to *fading*, which has a number of specific forms, and a variety of interesting solutions.

Normal Fading

We know that the ionosphere is not a solid object; it tends to drift around, wobble, and wrinkle. As we’ve discussed earlier, it’s rather amazing that it ever works at all, given all the parts that have to fall into place. So it stands to reason that a certain amount of fading is normal and expected.

If you have listened around the shortwave broadcast bands at all (and every self-respecting ham needs to do this occasionally) you will notice that even the strongest stations can have random dips in signal strength. These dips generally have a time scale of a few seconds to perhaps a minute or so. And they can be surprisingly deep. One of the primary incentives for the invention of the automatic gain control (AGC) or the earlier-named automatic volume control (AVC) on broadcast receivers was to compensate for ionospheric fading. Incidentally, skywave AM broadcast signals can have dramatic fading, as well. A typical plot of signal strength over a period of minutes (**Figure 13.1**) will show just how deep “normal” fading can be. This plot shows the signal strength of the 15 MHz WWV signal received in Fairbanks, Alaska, on a quiet day from 11 AM to 1 PM. Vertical scale is in relative S units to represent what you will typically see. Actual *field strength* variations are much deeper. The same data is shown with some mathematical smoothing for clarity in **Figure 13.2**.

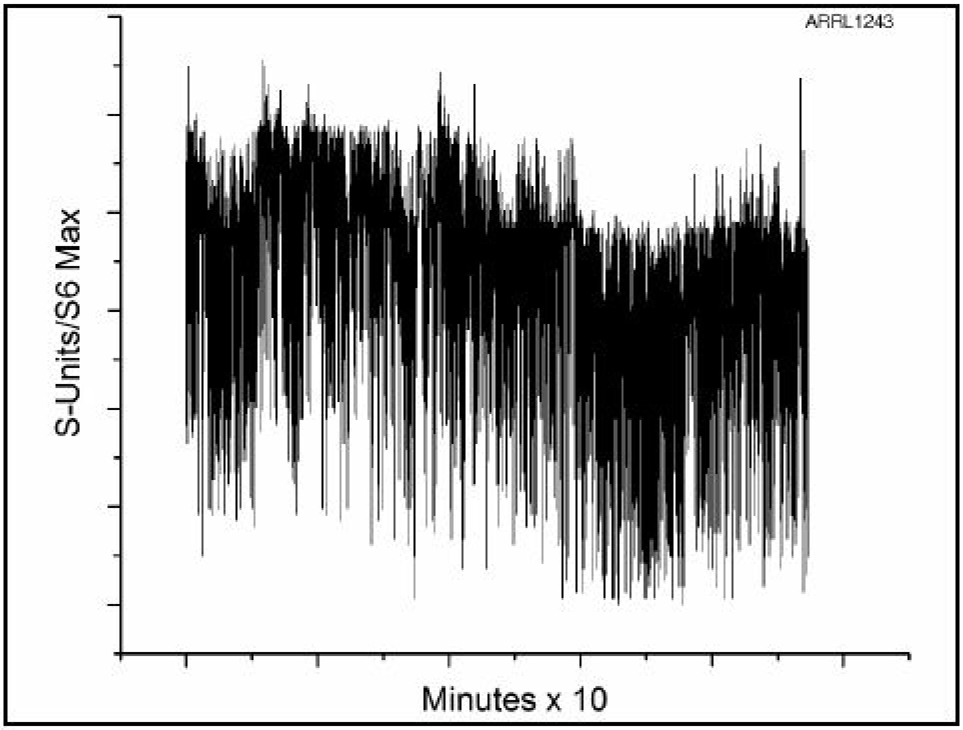


Figure 13.1 — Ionospheric fading can be surprisingly rapid and deep. This graph shows the signal strength of 15 MHz WWV over a period of two hours, monitored near Fairbanks, Alaska. Maximum signal strength is set at "S-6" at the distant receiver. Rapid fading is primarily caused by phase cancellation as reflected signals from different points of the "wobbling" ionosphere add and subtract in a random fashion. "Wrinkles" in the ionosphere can also simply redirect the reflected wave to different points on the Earth's surface.

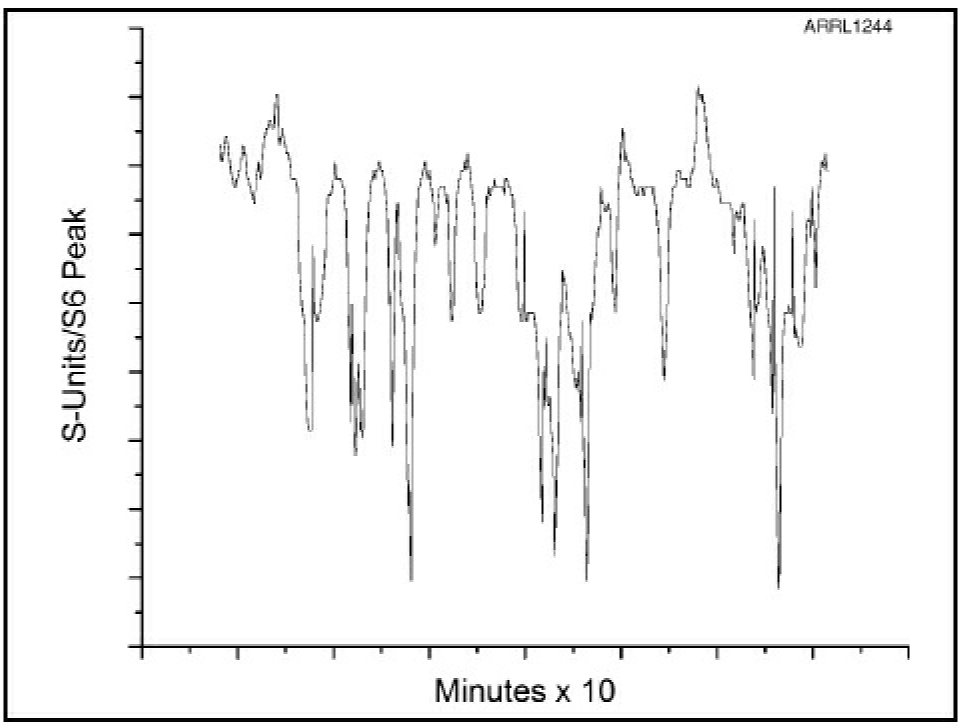


Figure 13.2 — This graph shows the same data as Figure 13.1 but with some mathematical averaging applied, so the peaks and valleys in signal strength can be seen more clearly.

Although localized pockets of absorption *can* occasionally cause fading, this is not the most common process. Typically, fading occurs when a signal is *redirected*, that is the reflecting surface that normally reflects the signal toward your receiver is "warped" and so reflects the signal in question somewhere else, causing it to skip over your head or make landfall before it gets to you. We also cannot discount *lateral* deflection, where the signal is redirected in azimuth due to the Earth's magnetic field. However, this is usually a long-term effect, not generally responsible for the short-term excursions we normally associate with fading.

In reality, it's a good thing that most HF radio transmissions are *not* laser beams; if such were the case, fading would be much more dramatic. As it is, radio reflections typically involve a large swath of ionosphere, so that localized ripples are somewhat averaged out. In other words, the ionosphere has somewhat of a *diversity* property built in, working in concert with a typically wide beam of an Amateur Radio transmission, or even a broadcast transmission.

Most radio signals take a number of paths to get from point A to point B, even if these multiple paths are very close to one another.

Phase Cancellation and Reinforcement

Another common form of fading is phase cancellation between a ground wave and a skywave, when both waves arrive at the receiver at the same time (**Figure 13.3**). You are most likely to notice this effect if you're listening to a relatively distant AM broadcast late at night while driving. You will often hear a station fade in and fade out at regular intervals of distance, often accompanied by large amounts of distortion during the "down-fade" intervals. The signals reinforce when they arrive in phase, and tend to cancel out when they arrive out of phase. Depending on the relative strengths of the skywave and ground wave signals, the cancellation can be nearly total.

AVC, AGC, and Their Kin

The traditional method of dealing with short term fading has been the AGC circuitry incorporated in nearly every modern receiver, which is also closely tied in with the typical S meter. Over the years, the AGC has evolved, as understanding of ionospheric fading (and improvement in radio circuitry) has progressed. If the attack time and recovery time of an AGC system are closely matched to the ionospheric fading characteristics, deep fading can be made almost unnoticeable in actual operation.

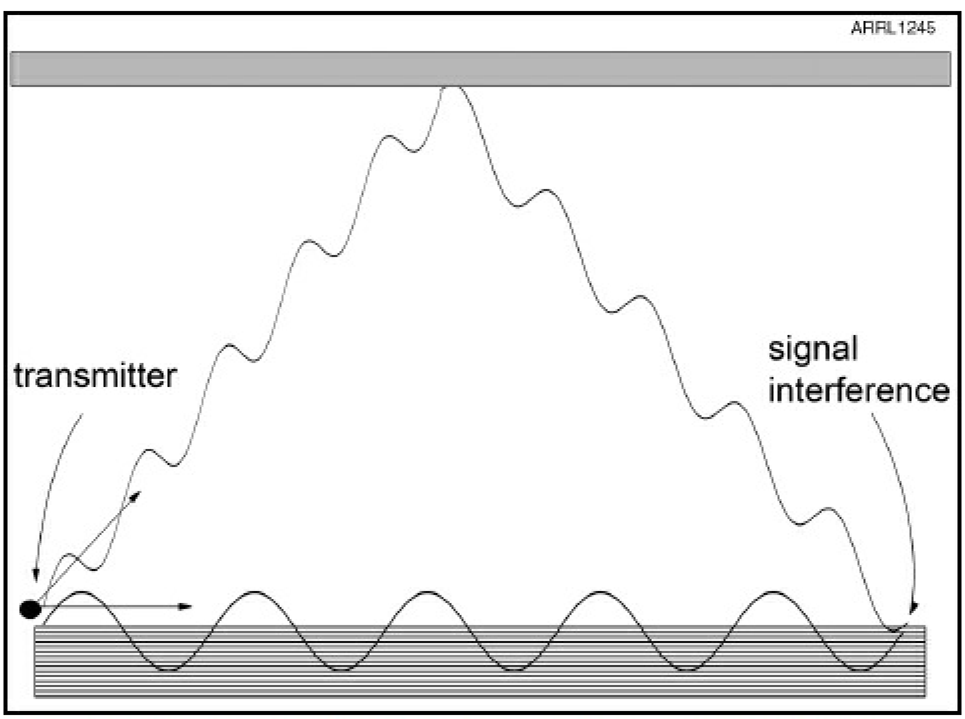


Figure 13.3 — Wave interference between a skywave and ground wave can be either constructive or destructive. Since the ground is stationary, and the ionosphere isn't, this can result in rapid fading over time.

Dispersion

We've established that the ionosphere is *dispersive,* that is the refractive index is dependent on frequency. Most hams understand this when they go about selecting a band to operate on. What most don't realize is that this dispersion can be significant within the bandwidth of a single radio transmission! If you are operating full AM, the upper sideband, being higher in frequency, will penetrate the ionosphere slightly further than the lower sideband, and therefore has a slightly longer propagation path. This is not to imply that if you transmit an AM signal, your lower sideband will arrive in San Francisco, and your upper sideband will arrive in Los Angeles. In most cases, both paths comfortably overlap. But not always. There are many cases of *selective fading* where two segments of one's transmission arrive far enough apart to not be receivable with the same antenna! In fact, this was one of the compelling incentives for the development of Single Sideband. An SSB signal has less frequency dispersion than a full AM signal, and is more likely to arrive intact at a distant location.

But you don't even have to be transmitting voice to experience selective fading issues. The problem first became…well…a *problem* during the 1930s, and was further more problematic when the first radioteletype (RTTY) circuits were deployed. RTTY uses frequency shift keying (FSK) to generate MARK or SPACE tones, and typically, these are just a few hundred hertz apart. As it turns out, the MARK and SPACE tones can show severe selective fading issues, which could make communications utterly impossible. And it is the same phenomenon that makes HF packet so unreliable; more advanced methods such as PACTOR take into account selective fading.

Diversity

This brings us to a topic and a method, which while being widely deployed in military and commercial applications, has seen limited use in Amateur Radio, namely *diversity reception.* Now, there are a few different types of diversity used in various radio services, but here we are speaking of *spatial diversity.* A diversity receiving system has two widely spaced antennas, two receivers, and a method of locking the receivers directly together in frequency. Often, they are connected together to share a common *detector* bus, as well.

The diversity method was developed by Harold H. Beverage (famous for the low noise receiving antenna bearing his name) and his colleague H.O. Peterson. A fascinating history of their discovery of the mechanism of selective fading can be found here: **www.radioblvd.com/DiversityDD1.html**.

Fundamentally, the solution to selective fading is to have two receivers far enough apart so that, at least statistically, neither one will fade out entirely at the same time. You may have noticed in the cited article, that Beverage and company had built a large number of *triple diversity* receiving sites. This may be a little overkill for amateur operation, but certainly worth experimenting with!

It should be fairly obvious that you might have to experiment a bit to find not only the optimum spacing for diversity receiving antennas, but also the best orientation relative to one another. And of course, this can only be optimum for one frequency, or perhaps a narrow band of frequencies.

Because of the complexity of a true diversity system, the place in Amateur Radio where it is most likely to be of great benefit would be in a fixed point digital link, such as those commonly set up for PACTOR III. The combination of selective fading immunity afforded by such protocols *in addition* to a diversity reception scheme would make for an incredibly robust and reliable HF link.

Synchronous Detection

While not a diversity method itself, so-called *synchronous detection* can reduces or eliminate several problems relating to selective fading. Remember our example of driving along listening to a distant AM station, and how you'd periodically receive the signal with lots of distortion? This distortion is a result of the *carrier wave* going away, either due to phase cancellation between the ground wave and the skywave, or simply severe selective fading.

The use of a *synchronous* detector mitigates the problem by supplying the missing carrier with a local oscillator. Synchronous detection has been around for a long time (in fact, the reception of suppressed carrier SSB is a form of synchronous detection), but it has only recently been used to any significant degree in Amateur Radio.

A true synchronous detector has a phase locked loop that is locked to the carrier frequency of an incoming AM signal, which is then used in a product detector to demodulate that AM signal. If the original carrier level drops below the level necessary to maintain phase lock, the free-running PLL is still stable enough to fill in until the carrier returns to a suitable level, thus effectively getting rid of the selective fading distortion problem.

The synchronous detection method should be of interest to anyone operating AM, whether a boat anchor aficionado, or any ham wanting to experiment with full AM. There are still many advantages to full AM, especially with regard to ionospheric research and experimentation. See the Appendix, Understanding AM for more information about this mode.

Frequency Diversity

While frequency diversity is a method generally used by commercial entities at microwave frequencies, it still has potential for HF. Radio amateurs instinctively use a form of frequency diversity whenever they change bands to accommodate changing conditions. At a somewhat higher degree of sophistication is *adaptive frequency selection*, more often used in military applications, where optimization of an HF frequency is determined by "pinging" a path and adjusting the frequency according to best propagation. It is questionable as to whether this method should ever be used on amateur bands, as it has the potential of causing tremendous interference if misapplied…but it does demonstrate a great truth. You can only determine the best path to a distant station by acquiring some kind of feedback *from* that distant station. The feedback can consist of something as simple as an honest signal report (you *always* give those out, don't you?). Or it can be something more sophisticated, such as an in-band telemetry signal from the other end (which is perfectly legal, by the way, though seldom if ever implemented).

One method that is tantamount to frequency diversity is *full duplex* operation. While standard practice on the maritime bands, it is essentially never used on amateur bands (except for some amateur satellite operation). As we have established, HF propagation is seldom reciprocal; the best frequency (or even band!) from point A to point B might not be the best frequency (or band) for going from point B to point A. With full duplex, you can optimize the AB and BA paths and equipment separately.

Full duplex SSB operation requires a bit more station complexity, of course, since you need two receivers and two transmitters. But for reliable emergency communications, it is well worth the effort, and the advantages are quite obvious, the primary one being there is no delay time between transmission and reception, a must for rapid-fire emergency and tactical communications. (By the way, this somewhat simplifies interfacing with a phone patch since there are no VOX adjustments to fiddle with, or PTT switching involved. And yes, phone patching even works with cell phones!) There's also the simple matter of redundancy; two radios are usually better than one. By the way, contrary to some "information" out there, full duplex on HF is perfectly legal; it's the same as running split. (While some hams argue that running split uses up more bandwidth, this is really a highly illogical conclusion).

Although space diversity antenna systems can use up a bit more real estate than "normal" antennas, receiving antennas can be much smaller than transmitting antennas in many cases. Space diversity can work with small tuned loops just as well as with full sized dipoles or other large antennas. It is an old, proven technique that deserves a place in Amateur Radio, especially in emergency communications scenarios.

Stretching Your Days

Certain antennas have a reputation for "working longer hours" than others. In other words, as daylight fades along with its attendant good propagation (at least on the high bands), some antennas seem to keep working later. Likewise, some antennas are credited with "waking up earlier" too. While the debate, informed as much by folklore and mythology as by science, promises to continue for the foreseeable future, the *principles* necessary for "long day" antennas are well known. Simple geometry tells us that as the daylight path approaches the horizon, an antenna with a lower angle of radiation will generally work better. And, as we've established, up to a certain point, the higher the antenna, the lower the angle of radiation, all other things being equal. One antenna that's often credited with having "longer days" is the cubical quad. Most likely this is due to the fact that, in comparison with a Yagi of equivalent gain, the quad's pattern is a little more "squashed" in the vertical plane. In addition, most hams, when talking about the height of their antennas, use the height of the feed point as the reference. A cubical quad fed at the bottom has a lot of activity happening well *above* that point, so in effect, the center of radiation is a bit higher than that of the Yagi. This demonstrates why it's so important to use uniform standards when comparing any two antennas. Most hams are guilty of comparing apples to oranges a lot of the time.

But there may be some validity to the cubical quad advantage in that the cubical quad occupies quite a bit more volume of sky than the Yagi. In other words, it probably has a bit of *built-in space diversity* due to its larger enclosed volume. It is unknown whether anyone has done any rigorous experimentation to prove this out, but it might be well worth the time for some ambitious hams to actually do this.

Multipath Matters

Multipath propagation, that is the ability for a single transmitted signal to take more than one path from Point A to Point B, can manifest itself on wavelengths anywhere from ELF up to the color blue and beyond. Whenever a multipath situation exists, there is always the potential for *constructive* or *destructive interference.* Of course, nothing is inherently wrong with either type of interference; all directional antennas rely on interference for their very function. But it's a different situation when two or more paths have a relative delay time on the order of thousandths or hundredths of a second instead of *microseconds.* Long time multipath effects can have detrimental effects on audio intelligibility or the ability to decode digital signals at all.

Most HF operators of long standing have heard both short-path and long-path DX signals, often both at the same time. When the conditions are such that both signals can be heard with approximately the same signal strength, the result can be anything from "curious" audio effects on phone to complete *blurring* or *smearing* of CW or digital signals. The most obvious solution to this is to have enough antenna front-to-back ratio to eliminate one of the signals. This may be a lot easier said than done, especially on low bands. Fortunately such conditions are fairly rare, enough so that they're generally more of a curiosity than a problem.

Multipath problems can be severe — as well as far more frequent — when using *near-vertical-incident-skywave* (NVIS) methods. This is because one can often receive both F1 and F2 layer reflections as well as both the X and O mode waves from each! This is more likely to happen near the magnetic equator, where the X and O mode signals are relatively well overlapped in azimuth, but rather divergent in reflection height. (In places like interior Alaska, the X and O mode signals are highly divergent in *both* reflection height and direction, so you don't often receive them both at the same place, especially in an NVIS configuration.) Note: the reason high latitude ionograms so clearly show the X and O traces is because the transmitting antennas are rather wide-beamed, so while being nominally "VIS" patterns, they cover a lot of sky. The same is essentially true of the receiving antennas as well, unless specially configured to be narrow-beamed. More on this in the upcoming chapter on NVIS methods.

So, what *can* be done to eliminate multipath if both signals are coming from the same direction, but delayed in time, such as in the case of simultaneous F1/F2 reflections? Obviously, no known antenna configuration can solve this problem. However, some intriguing possibilities exist in the area of digital signal processing. Since time-delayed multipath signals are generally very similar in *pattern* to the desired signals, DSP can be used effectively to erase the duplicate signal. This is not much different from using a DSP audio notch to get rid of an offending carrier; it just takes a little more processing and sophistication to get rid of a vocal snippet. DSP is evolving at an impressive rate, and as it is used to address these particular propagation issues, we can expect to see some significant improvements in performance, even if overall HF conditions deteriorate.

In Conclusion

There is a lot more experimentation that can be done with diversity receiving methods. Several commercial radio manufacturers have incorporated diversity reception methods into their HF transceivers. We have included a couple of block diagrams for effective diversity reception to get you started in an Appendix.

WWV and Channel Probes

One of the most useful tools for studying and predicting the ionosphere is free for the taking. We're talking about the continuous broadcasts of NIST stations WWV (Fort Collins, Colorado) and WWVH (Kauai, Hawaii). NIST, the National Institute of Standards and Technology, formerly the National Bureau of Standards (NBS), has continuously transmitted time, frequency, and propagation information since nearly the beginning of radio. There is a wealth of information in these broadcasts, both hidden and not-so-hidden, that is of great benefit to radio amateurs. You are strongly urged to investigate the NIST Time and Frequency Services website at **www.nist.gov/pml/div688/grp40/** to learn more about these stations and their broadcasts.

The repeating broadcast format, valid for all the WWV frequencies is shown in **Figure 14.1**. According to the website, beginning in April 2014 there is also an experimental temporary broadcast on 25 MHz, which may become a permanent feature in the future. For many years NBS broadcast on 25 MHz, but discontinued that broadcast around the time it became NIST. During periods of more solar activity, this 25 MHz broadcast is increasingly useful.

It is important to note that not only are the WWV and WWVH broadcast frequencies extremely precise, but so are the audio tones, as well as the *effective radiated power*. Most hams are well familiar with the "atomic accuracy" of WWV's time signals, but don't pay much attention to, and seldom use, the precise RF radiation characteristics of the station(s). NIST's scientists and engineers have put a tremendous amount of time and energy into producing this world class instrument, freely available to anyone with a general coverage receiver. Hams would do well to take advantage of all WWV has to offer.

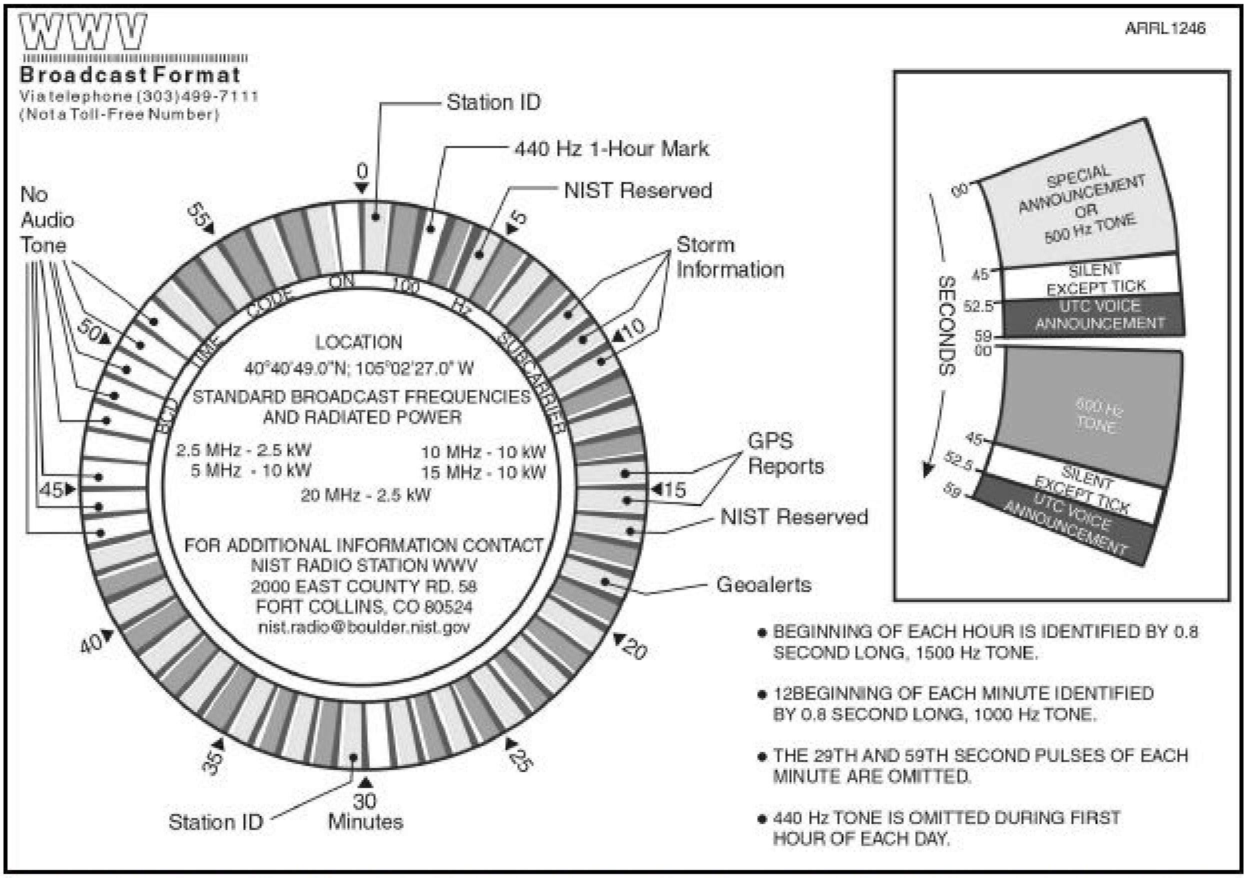


Figure 14.1 — The WWV schedule. The repeating broadcast format is valid for all the WWV frequencies.

In addition to the better known WWV and WWVH broadcasts, there are the VLF broadcasts of WWVB at 60 kHz, also originating from Fort Collins, Colorado. This is the broadcast that most low-cost "atomic accuracy" wall clocks use for their reference. It is a fun and somewhat challenging project to build a receiver for 60 kHz, and is good practice for any ham interested in experimenting with the "long waves."

More than Time

There is no excuse for any 21st century radio amateur to not know exactly *where* he is as well as *when* he is. Of course, most hams have set their station clocks to WWV for the better part of a century. The master clocks at WWV have always had atomic accuracy, and hams have, for quite some time, been able to come pretty close to that accuracy…at least for the short term. Only in recent years have hams (or anyone else) had access to equivalent accuracy when it comes to *position.*

Today, with the ready available of GPS receivers, any ham or other person can determine their absolute location on Earth to within a few feet. Needless to say, WWV knows the location of their antennas to equal accuracy.

"So what?" you may ask. Well, if we know the precise location of a transmitter and a receiver, say to within a few wavelengths, *and* the precise time at both locations, we can do some really interesting experiments.

One of the features of a WWV broadcast is the "time tick." These once-per-second, very precise audio tone bursts (which only sound like a tick to the human ear), give us the ability to do "time of flight" measurements. That means measuring the time it takes for a radio signal to get from point A to point B. Now, obviously, if we're using WWV as *time* reference by itself, it isn't too useful for determining how long it takes for a WWV tick to get from Fort Collins, Colorado, to wherever we happen to be. However, with GPS we can have an absolute, coordinated and synchronized time reference. With some fairly simple apparatus, namely a general coverage receiver, an oscilloscope, and a GPS receiver (optimized for time), we can now look at the absolute time of travel of the leading edge of WWV's tick to our location, easily to within a few microseconds.

As just one example, with this information we can determine with a lot more reliability how many hops it takes for a signal from WWV to get to us. Or, perhaps we already know, by some other means, just how many hops it takes to get to us. In this case, we can determine the reflection height, again by doing a time of flight measurement.

Knowing the total number of hops as well as the height per hop, allows us to determine the total path length. Well, this probably brings up another "so what?" which we will address immediately.

Absolutely

One of the reasons hams are careless about S meter reports is that signal strength is "all relative, anyway." All that's really important is how loud the station is compared to the previous station, right? Ionospheric propagation is random and nobody knows what their antenna gain is anyway, so what good are S meter readings? These are the kinds of comments one always hears implied, if not stated explicitly on the air.

The truth is, absolute measurements *are* absolutely possible. Would WWV spend so much effort assuring absolute ERP radiation specifications if they weren't?

Let's make this very simple. If you know a distant transmitter's absolute effective radiated power (ERP), you know the distance this radio signal travels to get to you, and you know your antenna gain and receiver sensitivity, you can *absolutely* determine the path loss between the distant station and you. Or, conversely, if you know the path loss and your antenna gain and receiver sensitivity, you can determine *absolutely* the transmitting station's effective radiated power!

The bottom line is, we radio amateurs already have at our disposal the means to make accurate, absolute measurements of just about anything that needs measuring — if we simply apply a little scientific discipline. Of course, there's more to Amateur Radio than making scientific measurements; most of us are happy to just make communications happen. But a large number of hams also want to make their signals better, and yet go about it in a very haphazard manner.

Probing for More Information

The channel probe, described in Chapter 9, consists of a beacon transmitter paired with a dedicated receiver, the purpose of which is to continuously monitor a propagation path between the two points. The nice thing about WWV is that the amateur community can put together an extensive network of channel probes without having to pay for a single transmitter.

Signal strength is just one piece of useful information you can derive from a channel probe, however. Measuring Doppler shift is a great way to determine the stability of the ionosphere, and a sensitive predictor of disruptions on the horizon. Although it's *easiest* to measure Doppler shift when you have a single reference oscillator for both the transmitter and receiver, as in the case of an ionosonde, most modern amateur receivers have more than enough stability to detect Doppler shift of incoming signals.

If you have a computer with a sound card, all you need to do is connect your receiver's audio output to the sound card line input, and tune the receiver, in USB mode for a low frequency beat note (100 Hz or so, high enough to be in the passband of your sound card) between the BFO and the carrier of the station in question (WWV in this case). You may have to use your IF shift to scoot your SSB filter down to accommodate the tone, if it normally cuts off at 300 Hz or so. Doppler shift will manifest itself as a slight audio frequency shift, upward in frequency for an advancing (lowering) ionosphere, and lower in frequency for a retreating (rising) ionosphere. If you have an audio program such as *Audacity*, you can do an FFT frequency measurement of the beat note. Or if you want to go low tech, you can just connect a frequency counter to your receiver's audio output.

Minding the Gaps

Since no WWV transmission is inside any amateur band (the closest it gets to being so is on 10 MHz, which is just a sliver below the 30 meter band), you may have to interpret the results you get by listening to WWV to actual Amateur Radio conditions. In most cases, little extrapolation is necessary. The 15 MHz WWV signal is a pretty doggone good indicator of 20 meter propagation. The 20 MHz WWV signal is an even better predictor of 15 meters.

On the other hand, there's a pretty big gap between 5 MHz and 10 MHz, and conditions can vary radically between these two frequencies. Which of these frequencies is a good predictor of 40 meter propagation? This is a real tough call. Fortunately (or unfortunately) there are usually a lot of shortwave broadcast stations in this gap, and they are *very* reliable indicators of 40 meter propagation.

For the newcomer who might feel chagrined by the number of high power shortwave stations we sometimes have to work around on 40 meters between 7200 and 7300 kHz, remember this. The broadcasters are there specifically because it is such a good segment of HF real estate. There are more hours per year — indeed, per decade — where propagation on 40 meters is good to *somewhere*. This is why for the better part of a century 40 meters has been the *default* HF band. In fact, a recent semi-scientific Internet poll asked the question, "If you had only one HF Amateur Radio band, which one would you choose?" Forty meters won, hands down. So, despite all the warts and wrinkles and frustrations, 40 meters is a wonderful band. And it's also arguable that it's wonderful for more hours of the year than any other band. It is perhaps the only true all-around day-and-night HF band.

Well, back to the gaps. Since it's very close to WWV's 5 MHz broadcast, we can expect that WWV is a great predictor of 60 meter propagation. Incidentally, as limited as the current 60 meter allocation is, it fills a much needed gap in propagation between 80 and 40 meters. A small but growing contingent especially appreciates this band in interior Alaska, where it has nearly ideal NVIS properties for the region. It's a resource more hams should use.

Our 80 meter band falls into a pretty big one, too, between the WWV 2.5 MHz and 5 MHz broadcasts. The 80 meter band is an interesting case in itself. Percentage-wise, it's a very large band. For a long time, 80 meters has been "psychologically" divided into 75/80 meters, primarily because of the somewhat different "missions" of the upper and lower parts. But few hams realize the profound difference in propagation that can exist between the top end and the bottom end of the 75/80 meter band…certainly enough to deem the band *physically* worthy of being two different bands.

Long distance 80 meter propagation resides in that thin blue line (almost literally!) between the highly absorptive D layer and the lower edge of the F region. There can be remarkably abrupt changes in the amount of absorption as you move between 75 and 80 meters. Some hams report that it's just plain easier to work DX on 75 meters than 80. This is despite the fact that 80 meters is populated by a lot more weak signal modes.

At any rate, 5 MHz WWV is probably a weak indicator of propagation on 75 meters, and a pretty poor one for 80 meters. Which brings us to, last if not least, 2.5 MHz. Although 2.5 MHz is the closest WWV frequency to 80 meters, it's not close enough to really do much good. (At least for amateurs. For shortwave listeners, there's plenty of commercial "tropical band" broadcasting that 2.5 MHz WWV does predict well.) The bottom end of 80 meters is still 1.4 times the frequency of WWV. That's a big gap. So, probably the best predictor of 80 meter propagation is…80 meters!

What about 160 meters? Likewise, it's also a big band, with significant differences in propagation between the top and bottom ends. As it turns out, however, the best predictor of 160 meter propagation is AM broadcast DX, particularly the top end. When you start hearing lots of distant stations on your car radio, it's generally time to fire up the 160 meter station. Keep in mind that the AM band is absolutely *huge*, percentage wise, with the top end being well over three times the frequency of the bottom end. So be sure to use the upper end of the dial as your "160 beacon."

But What about Me?

Needless to say, Fort Collins, Colorado, where WWV resides, is not the center of the radio universe, and so, may not be the best predictor of radio "goodness" in your bailiwick. Even the fine folks at NIST realize this, which is why they duplicated WWV (for the most part) in Hawaii, in the form of WWVH. Ideally, to create the best overall "picture" it would be nice if WWVH were diametrically opposite the Earth from WWV, but since there's basically nobody *there*, they had to come up with a compromise. And WWVH is a good one. Between WWV and WWVH, you really can get a pretty fair overall global propagation tool. (Note of curiosity. Directly antipodal to Fort Collins, Colorado, is Amsterdam Island, a rock in the southern Indian Ocean. Earlier this year there was an actual Amateur Radio DXpedition to Amsterdam Island, FT5ZM. Just goes to show you that radio amateurs will do just about anything to check out radio propagation theories.)

Triangulation in 3D

One of the most useful features of a channel probe is that you can do some very interesting path studies. By now we trust that we have driven home the concept that radio waves (or *rays*) are affected both in azimuth and elevation by a great number of factors. In the case of a channel probe, presumably we know the location of both the receiver and the transmitter in question. Knowing both ends of the "circuit" we can make some fairly reliable determinations about the paths from point A to point B. However, this is not so straightforward in many cases.

Most hams at one time or another have been involved in fox hunts or hidden transmitter hunts. Locating a remote transmitter by triangulation is quite straightforward, if the radio signal travels unperturbed in more or less a straight line from the transmitter to the receiver. Some very simple trigonometry will allow you to pinpoint the location of the transmitter by taking two directional "fixes" from widely separated locations. The farther these points are from each other, generally the more accurate your prediction will be. As you get closer to the transmitter, you can use simple signal strength to do the final homing in on the target.

This process is a whole lot more interesting if the propagation path is via the ionosphere. Before 2 meters became immensely popular, many ham radio clubs put on 80 meter fox hunts, which were far more challenging even under the best of conditions.

It is relatively easy to triangulate on an 80 meter signal if it is primarily on ground wave. But what if the path is not ground wave, but rather skywave…or a combination of both? Can it even be done? Yes it can, but it takes some clever techniques.

One of the more practical uses for HF direction finding is determining the location of interfering or interloping HF stations. The ARRL Intruder Watch program is ever on the alert for foreign broadcasters and other entities that creep into the amateur bands uninvited, and generally unappreciated.

It's a relatively simple matter to determine the general direction in azimuth using a directional antenna — assuming of course, we take into account all the X and O propagation information discussed in previous chapters. But finding the distance to the offending station (or interference source) can be extremely challenging if the distance is great (such as half way around the world).

To achieve any kind of accuracy using two-point triangulation requires that the receiving stations are relatively far apart compared to the distance to the transmitting station, which may be impossible to do from the same continent. But the problem is even more "tricky" if the received signal is arriving from a high angle, which also suggests that it has probably taken multiple hops to get to "us." Without some very clever methodology, the best we can do is pinpoint the location in the ionosphere where the *last* hop occurs. But in many cases, this may be enough, if we make some reasonable assumptions.

Vertical Plane Triangulation

As suggested above, triangulation in the azimuth is familiar methodology for most hams, Triangulation in the vertical plane is rarely employed, and yet quite important to fully understand a lot of ionospheric phenomena. Very rough estimation of vertical angle of arrival can be done with a single antenna or array, but to accurately pinpoint of a radio source (which may be a reflection or *secondary radio source*) generally requires two widely separated receiving stations, each at a significantly different distance from the source. See **Figure 14.2**. If each of the stations can determine the angle to the source, then simple trigonometry can be used to determine the height of the reflection. Once this reflection point is determined, then one can make some reasonable assumptions as to the location of the primary transmitter.

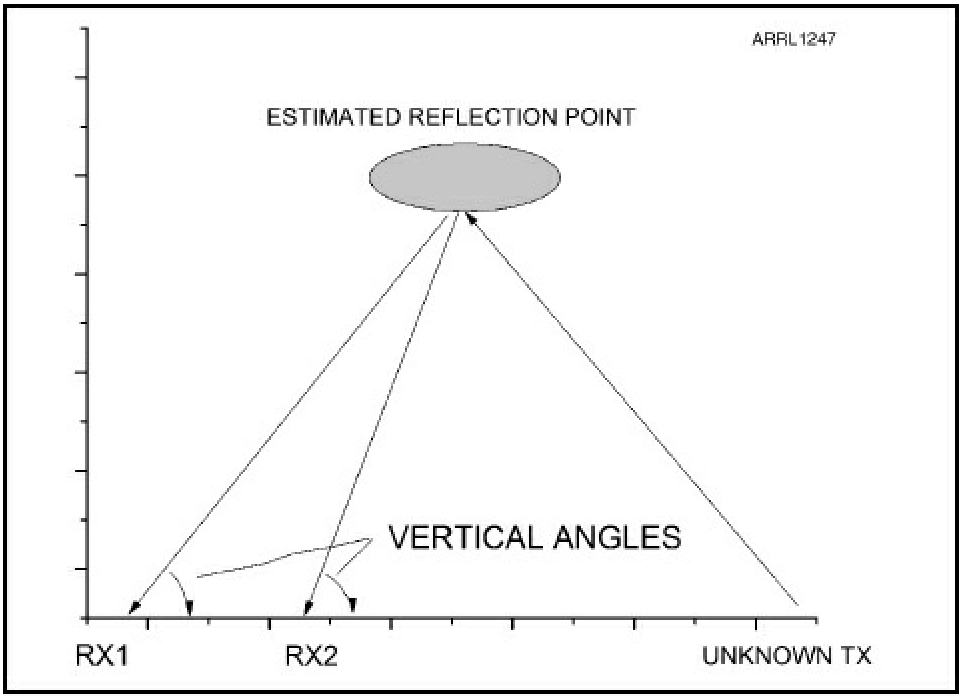


Figure 14.2 — This illustration shows the rarely used method of *vertical angle* triangulation. The principle is identical to triangulating in azimuth, but can yield far more meaningful results for the student of the ionosphere. The farther the two receive sites are separated, the greater the accuracy of the measurements.

The Interferometer

The *interferometer* is an instrument that is probably worthy of an entire chapter, except for the fact that most radio amateurs probably don't have the real estate to build a really effective one, at least for HF frequencies. Instead, we will discuss the principles of interferometry in just enough detail that you can adapt it to your needs and curiosity level.

Any directional antenna — and we truly mean *any* directional antenna — ultimately functions on the principle of wave interference. This is true whether you're working with a Yagi, a phased array, a parabolic dish, or even a random wire. Any time you have multiple sources of radiation, there is the certainty of either constructive (in phase) or destructive (out of phase) interference at some distant location. Perhaps you might ask how a Yagi could be multiple sources of radiation. If you recall our strong emphasis on *reradiation* early on in this book, you will realize that every "reflector" is actually a secondary emitter. In the case of a Yagi, the driven element and parasitic *sources* are quite close together. The same is the case for the typical *phased array*.

The interferometer is, simply stated, a receiving array where the currents and phases on each of the elements is closely measured. There is no strict line of demarcation between a normal directional receiving antenna and an interferometer. However, it can be said that in an interferometer the elements are far enough apart so that there is no interaction (magnetic coupling) between the elements. A *large baseline* interferometer is one in which the spacing between elements is many hundreds or thousands of wavelengths apart at the frequency of operation. In such a case, each element is associated with a separate receiver.

Large baseline and very large baseline arrays are used extensively in radio astronomy. In order to reassemble a proper picture, the relative phases of each of the elements must be precisely synchronized. At one time this was an unthinkable prospect for any radio amateur; nowadays, fully phase coherent multiple receiving stations are within the reach of any technically astute radio amateur, at least into the HF frequency range. However, this is somewhat radio geeky stuff that is beyond the scope of this chapter, though you should at least know the technology exists.

A much more practical interferometer for most hams is a relatively widely spaced array of a few elements feeding a single receiver or a group of co-located receivers.

An interferometer array can be one-dimensional, such as a collinear array of dipoles. It can also be an end-fire array of dipoles. A two-dimensional interferometer can consist of a combination of collinear and end-fire dipoles. Theoretically, a 3D interferometer could be built of multiple layers of 2D arrays, but this is normally too unwieldy to consider for amateur work.

A wide-spaced square array of four horizontal dipoles can be used to determine both azimuth and elevation angle of incoming radio signals. Actually this can be done with just *three* dipoles in a triangular array, but the number crunching is a bit more involved.

The Direct Conversion Shortcut

The most elaborate array of antennas won't tell you much unless you have a receiver, or several receivers. In an interferometer, the receiver at the very minimum must extract signal phase information from its associated antenna element. In some cases, it's helpful to have amplitude information, but for strictly direction finding, this isn't too crucial. The relative phases of each element need to be accurately determined as well, which requires fully phase coherent receivers. This used to be a considerable engineering challenge, but not with the direct conversion receiver. Using simple direct conversion techniques, actually nothing more than a stable local oscillator and a number of doubly balanced mixers, a phase coherent receiver with any number of channels can be assembled very easily.

The output voltage of a doubly balanced mixer is a dc voltage which is directly proportional to the phase angle between the incoming signal and the local oscillator (which is on the same frequency as the signal of interest). By looking at the relative dc voltages coming from each of the balanced mixers, we can then determine the direction of arrival of a signal. The actual circuit configuration is covered in more detail in *Radio Science for the Radio Amateur* for those interested in pursuing this technology further.

Software and Other Tools

There are a number of extremely useful and fascinating propagation prediction tools available to the radio amateur. Often these are associated with antenna modeling programs, which makes perfect sense. Antennas are our only interface between radio propagation and everything else we do in the ham shack. We take advantage of propagation characteristics primarily by making adjustments to our antenna systems…ranging from minor tweaks to major overhauls.

One must be cautious when using propagation modeling tools, as with any computer modeling software. Computers are notoriously lacking in common sense; they won't actually think for you. You need to have a fairly good idea of what the answer will be before you feed your problem to a computer.

As we have already elaborated, only in very recent years have any propagation modeling programs (at least those available to the radio amateur) incorporated the effects of the Earth's magnetic field. Even now, you will be hard-pressed to find a propagation program that distinguishes between X and O modes. Nevertheless, armed with the software we do have, *and* considering the information we know about the mechanics of ionospheric propagation, we can do some pretty useful modeling. Hopefully, the "code poets" amongst us will be motivated to work more X and O parameters into propagation programs.

To get a glimpse of what's available in terms of propagation software, here is a great site: **www.dxzone.com/catalog/Software/Propagation/**. We'll take a look at a number of these programs and give you some pros and cons of each of them. Another good site is **www.astrosurf.com/luxorion/qsl-review-propagation-software-dos.htm**, which also includes some comprehensive reviews of existing propagation software. And there is a lot of information on the ARRL website at **www.arrl.org/propagation-of-rf-signals**.

Probably the first propagation prediction program (say that three times fast!) widely known to radio amateurs was *MINIMUF*. It was presented in December 1982 *QST* by Robert B. Rose, K6GKU. Written in a mere 80 lines of BASIC code, its purpose was to predict the maximum usable frequency (MUF) between two selected locations on Earth. Despite its primitive command-line coding, it served well for over a decade, at a time when hams were first inviting these strange devices known as computers into their ham shacks. Development of this software emerged from the Naval Ocean Systems Center (NOSC) as *MINIMUF 3* (about the same time as *MININEC 3* antenna modeling software came out). It certainly says something that the best and brightest minds on the planet could write a scientific computer program with 80 lines of code, while a typical modern operating system needs a million times that just to clear its throat!

Maximum usable frequency, as described earlier, is still one of the best benchmarks for predicting useful radio paths and times of operation. As a good rule of thumb, the lowest path loss occurs when the operating frequency is about 10% below the MUF. Obviously, you don't always have the option of selecting the optimum operating frequency based on MUF, since that frequency is as likely to fall outside an amateur band as within it. But MUF is *certainly* capable of telling you *which band* you need to be operating for the best results.

We can get a first order approximation of MUF from the simple formula:

$$\text{MUF} = \frac{\text{Critical Frequency}}{\text{Cos}(\theta)\ \text{Angle of Incidence}}$$

where the angle of incidence is the angle from the zenith (not elevation angle).

Of course we need to know what the critical frequency is first at the "mid-hop" which can only be measured with something like a vertical incidence ionosonde. *MINIMUF* was the first "ham-available" program that could derive the critical frequency directly from solar data. Of course you needed to get that solar flux data from somewhere, such as the WWV broadcasts.

As groundbreaking as *MINIMUF* was, it wasn't long before a number of hams ran across some deficiencies of varying severity in the modeling, especially with regard to polar paths. A search through the *QST* archives (available from the ARRL website to any ARRL member) for *MINIMUF* will reveal some interesting revisions and comments pertinent to *MINIMUF*. At this point in time, *MINIMUF* is probably only important as a historical curiosity; however, it's probably safe to say that all modern propagation prediction software has its roots in *MINIMUF*.

A fairly direct descendent of *MINIMUF* is *MicroMUF*, which was also written in BASIC. It was developed at BBC and Radio Nederland, primarily for use by avid European SWLs (shortwave listeners), but worked itself into the Amateur Radio ranks, where it is still highly popular. It lacks the elegant user interface of a lot of modern programs, but it is a powerful piece of software that addresses many of the issues with the original *MINIMUF*. As with *MINIMUF*, the source code for *MicroMUF* is readily available, an important consideration for the radio amateur who actually wants to continue to improve the software. Several third party user interfaces have been created to use *MicroMUF* as the "back end" as well. In this regard, the development path of *MicroMUF* has been quite similar to the many antenna modeling programs based on *NEC-2* and *MININEC*…that is, the addition of increasingly user friendly graphical interfaces to a "back end" that is already powerful and proven. (More information on antenna modeling may be found at **www.arrl.org/antenna-modeling**.)

Suite Deal

The Institute for Telecommunications Sciences (ITS) has developed an extensive suite of propagation prediction software in a package called *ITS-HF*. (Like nearly all HF propagation programs, this software is free). *ITS-HF* is also included as a plug-in with *4nec2* antenna modeling software.

At the core of *ITS-HF* are a couple of mature programs, *IONCAP* and *VOACAP*. *IONCAP* stands for Ionospheric Communications and Prediction program, while *VOACAP* was the same program with enhancements developed by Voice of America (VOA). A detailed history of the development of VOACAP is available here: **www.astrosurf.com/luxorion/qsl-soft-voacap.htm**.

Like a lot of hard core number-crunching programs, it was written in Fortran, but parts of it have since been ported to some other more modern languages. However, Fortran is still the benchmark program for most numerically intensive number crunching. It will be around for a long time to come.

At any rate, the *ITS-HF* suite allows you to insert just about every possible variable between a transmitter and a receiver, including such parameters as antenna gain or local noise levels. Then, through a series of subroutines, based on your priorities, it will tell you the total signal path, the total path loss, the loss per hop, the optimum working frequency, optimum beam headings for transmitter *and* receiver, optimum time of day, month, year, or even solar cycle. There are too many parameters to describe in detail here, but it only takes a few minutes to download and install the entire suite, so the best advice is to just start playing with the program. The documentation is extensive, but most of the input fields are pretty self-explanatory anyway.

One interesting feature of *IONCAP* is that it correctly assumes that there will be a number of possible paths a signal can take through the ionosphere between Point A and Point B. It allows you to choose *which* ionospheric layer you want to use…or all of them. This is especially useful for "The Magic Band" (6 meters), for instance, where sporadic E, F1, and F2 paths might exist simultaneously.

Proplab Pro, Version 3

Proplab Pro, Version 3 software (**www.spacew.com/proplab/**) deserves special note, because at this point in time, it is the only commercially available software (that we know of) that models both X and O propagation modes independently (or at all). It also explicitly calculates non-Great Circle paths and other parameters uniquely related to X and O propagation.

These features alone make it worth the somewhat hefty price (on the order of $350), for the serious radio propagation student. It also has some eye catching ray-tracing graphics, if that sort of thing is important to you. Of course it is; if you use this program at all, you'll find yourself using it a *lot.* You don't want to bore your eyeballs with lame graphics! See **Figures 15.1** through **Figure 15.3** for a glimpse of *Proplab Pro*'s capabilities.

Gray Area

Most worthy propagation prediction programs include gray line information (**Figure 15.4**), although it may be a bit difficult to call the presentation of gray line information "prediction." This is because the gray line is as predictable as clockwork; it's an astronomical phenomenon. There are a number of great free-standing gray line clocks and gray line overlays for propagation charts.

Just because the gray line is so predictable, doesn't mean it's of any less importance in radio propagation studies. In fact, all propagation prediction software requires that you input your time of day, or the other station's time of day, or both.

One aspect of the gray line that is not a gray area is its importance for any serious DXer. Many hard-core DXers confine their entire operation to gray-line contacts because this has the greatest possibility of yielding the best DX results most of the time.

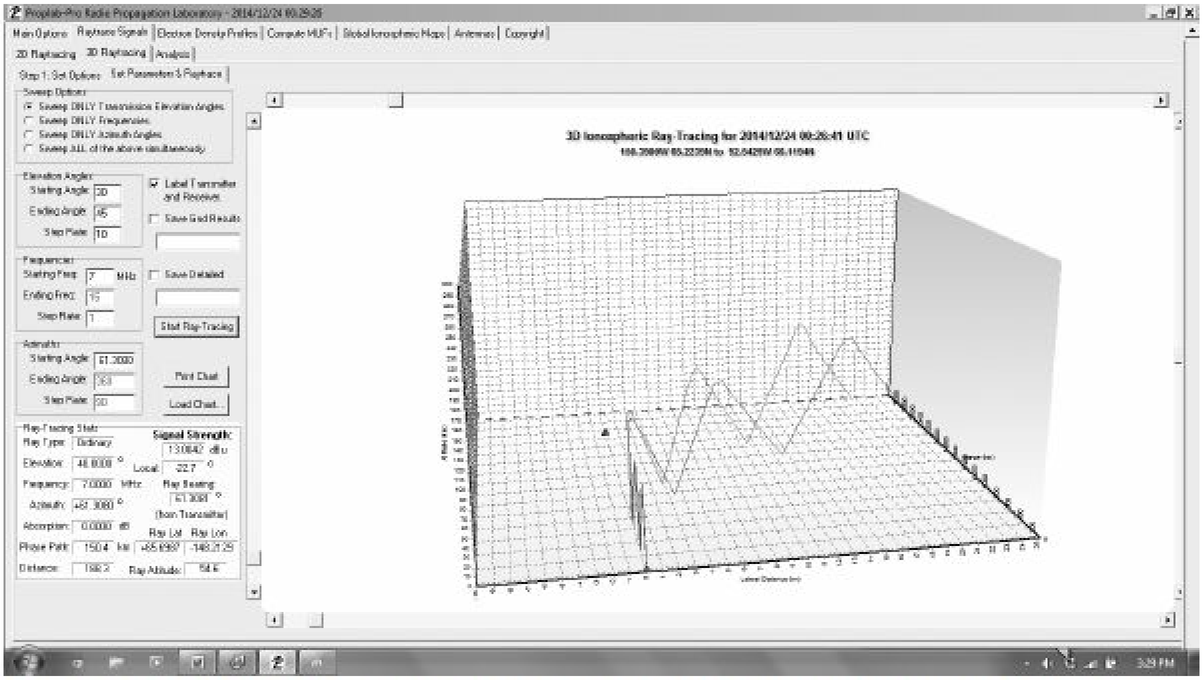


Figure 15.1 — *Proplab Pro 3* is a true ray tracing program showing the paths of both the O ray (in red on the computer screen, the top trace in this illustration) and the X ray (in green on the screen, the lower trace here). You can see all the hops between the transmitter and the receiver....if the signal can actually make it there. This plot shows a 7 MHz path between Fairbanks, Alaska, and Baffin Island, Canada.

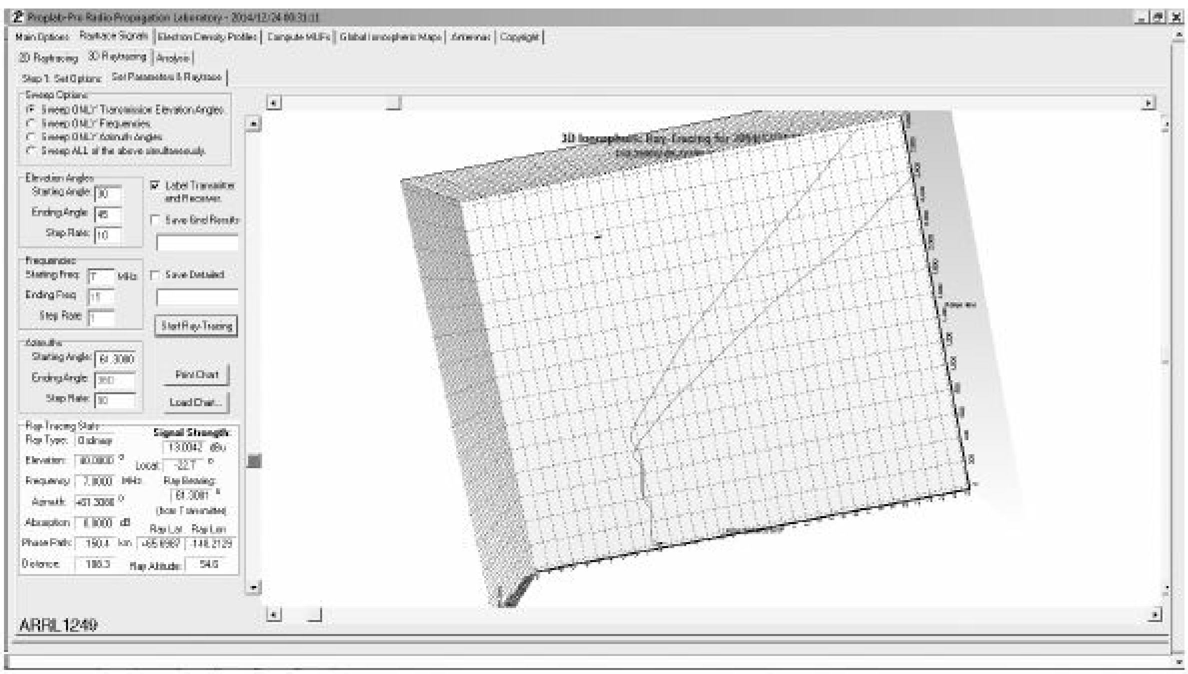


Figure 15.2 — Here we have the same source and destination plotted as in Figure 15.1, but with the graph turned so as to look straight down. Fairbanks, Alaska, is the small black square, and Baffin Island is the small triangle in the upper left. A straight line between these symbols represents a Great Circle path. Notice how far the rays diverge not only from the Great Circle route, but also from each other! *Proplab Pro* is the only propagation program available that accounts for the Earth's Magnetic field. (After I discovered this program, it's the only one I use anymore!)

Of course, this does not work everywhere. A good example of the "gray line exception" is here in northern Alaska, where in the dead of winter, the gray line is perfectly horizontal where it intersects Fairbanks. To shoot the gray line you need to aim due East or due West. If you aim West, you might graze some Eastern Europe, but if you aim due East, there's nothing there! The gray line goes into a black hole, ham radio wise. There are a number of other locations where no two distant locations (at least with any population) happen to lie on the same gray line. So, as useful as the gray line is for a lot of hams, it can be useless for a few others.

One clear advantage of working gray line, when possible, is that the gray line path is less frequency selective than other paths. Since it is the twilight zone between daylight propagation and nighttime propagation, it can be characterized as neither a "high-band" nor "low-band" mode. In fact, it is both. You have a much greater chance of earning 5-band (or more) DXCC "wallpaper" by working gray line than you do working "normal" paths.

Another advantage, from a software point of view, is that multi-hop propagation "number crunching" is a bit simpler. The reason for this should be fairly clear. Along the gray line, the ionization levels, reflection heights, and critical frequencies are going to be very similar. Therefore a single hop profile can be more readily extrapolated to a multi-hop path. Any point along the gray line is, from a propagation standpoint, the same time of day. This also holds true for any point on a line *parallel* to the gray line. Or, put another way, you don't have to shoot down the center of the gray line (which is not a line, but a rather fat band); you can hug one edge of it, as long as the station at the other end hugs the same edge. This is not to imply that you can't "cross" the gray line with great effectiveness; it's just that if you want to see the uniformity of gray line propagation, you need to run parallel to it.

This is another instance of why the "sloppiness" of a typical HF antenna pattern is actually helpful. Any practical ham radio beam pattern is going to effectively cover the entire width of the gray line, under most circumstances. Gray line propagation is, for the most part, a very *forgiving* mode of operation, one that works very well for most hams of limited means. The only real price you pay for being a dedicated gray line operator is sleep deprivation! No matter where on the gray line you happen to be, whenever there's decent propagation to where you want to propagate to, it always seems like it's three in the morning. Of course, there's no scientific proof for this!

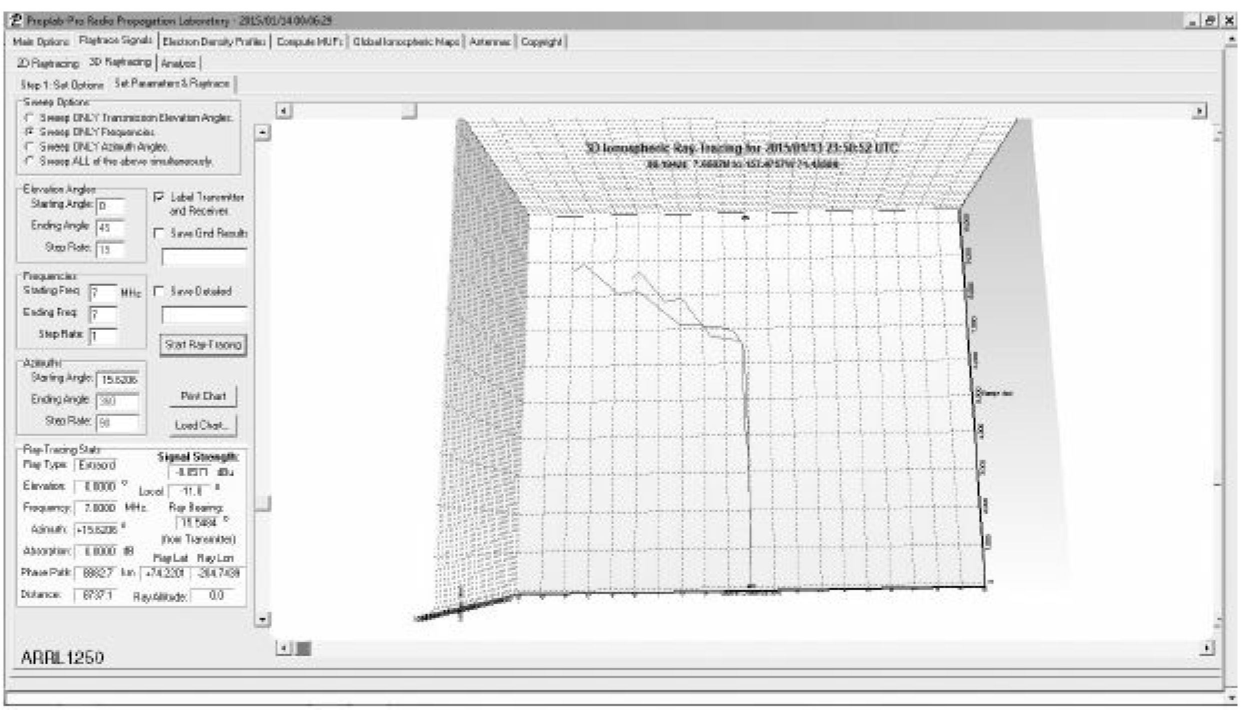


Figure 15.3 — Again from *Proplab Pro 3*, this illustration shows the divergence of the O (the trace starting furthest to the left) and X rays from Barrow, Alaska, to Sri Lanka on 14 MHz. Great circle path is from the little square at bottom center, and the top center triangle is Sri Lanka. The rays never make it.

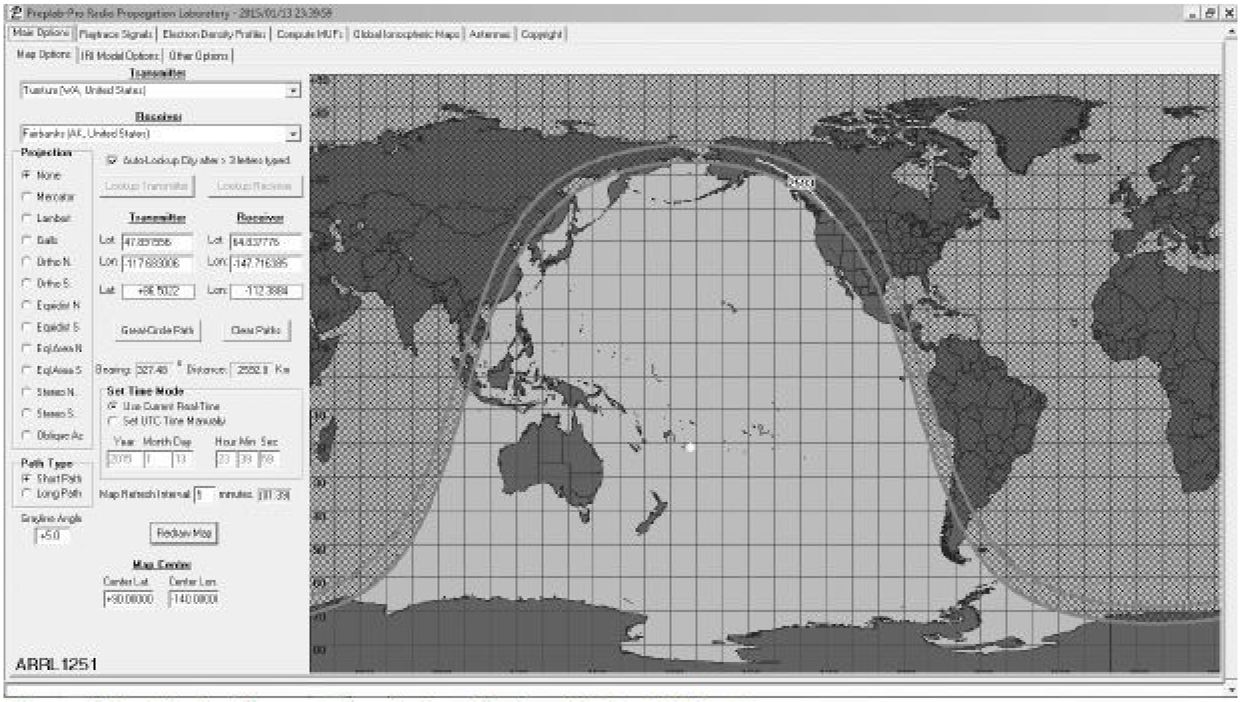


Figure 15.4 — The familiar gray line chart as displayed in *Proplab Pro 3*.

Long Path, Wrong Path

It's hard to talk about gray line propagation for very long without entering the discussion of long-path propagation. If you look at a gray line, you will notice it's not really a line, but a closed loop. There are always, at least in theory, two ways to get from Point A to Point B along the gray line. These two paths are known as short path, and long path. Now the obvious question arises, is one path necessarily longer than the other? What's the deal?

As it turns out, the "deal" is pretty much just statistical. In North America, at least, the two possible gray line paths to nearly any DX station will be radically different in length. This is largely due to the fact that 90% of the Earth's population lives above the equator, making things rather unsymmetrical to start with. (Since hams are, for the most part, a subset of humans, it's probably safe to say that 90% of the ham population lives above the equator, as well.) Now, any experienced DXer will tell you that the gray line is *not* the only place where long path can happen. It's just, in general, a lot *easier* to do long path via the gray line.

Now, to the best of our knowledge, there are no specific awards for working a DX station long path. There's no "DXCC short+long" award, last we checked (though maybe there should be). But working a station long and short path is a great deal of fun anyway, and at the very least it helps confirm the notion that the Earth is actually round, a fact of which hams need occasional reminding.

Beaming with Joy

Some of the earliest software used by radio amateurs was used for determining beam headings. Like the gray line map, determining beam headings is not really a *prediction* task, but on the other hand, it's not exactly a trivial pursuit, mathematically, either. If the Earth were still flat, determining beam headings *would* be a trivial matter, but thanks to Christopher Columbus' ill-conceived experiment, the Earth is now round, making the proper determination of a DX beam heading much more of a chore.

Fortunately for all involved, there are a number of programs out there that will give you correct beam headings to any place on Earth based on your location. You can order azimuth charts centered on your location from a number of sources as well (for example, online at **cqmaps.myshopify.com**).

If you want to go really old school, you can obtain a world globe, paste a compass card onto the globe, centered on your location, oriented so the NORTH is aimed at the North Pole. Take a piece of string, and run it between the center of the compass card and the DX location in question. Where the string crosses the compass card is the direction you want to steer your beam. This method is not as Neanderthal as it may seem; such readings can be done extremely quickly with some practice.

If you're like most hams, you'll probably want to use some technology from the current millennium. Beam heading software is in abundant supply. There are even logging and DX software packages, such as *Ham Radio Deluxe*, **www.ham-radio-deluxe.com**, that will do your beam steering for you with the appropriate interface.

Scientific Plotting Software

A picture's worth a thousand words. While not for the rank beginner, scientific plotting programs such as *Matlab*, *Mathcad*, *Octave*, and *OriginLab* can produce some amazing 3D images to help visualize a lot of obscure electromagnetic principles. **Figure 15.5**, a plot generated with *OriginLab 9.1*, shows the relative field strength versus azimuth and elevation of an inverted V antenna over real ground.

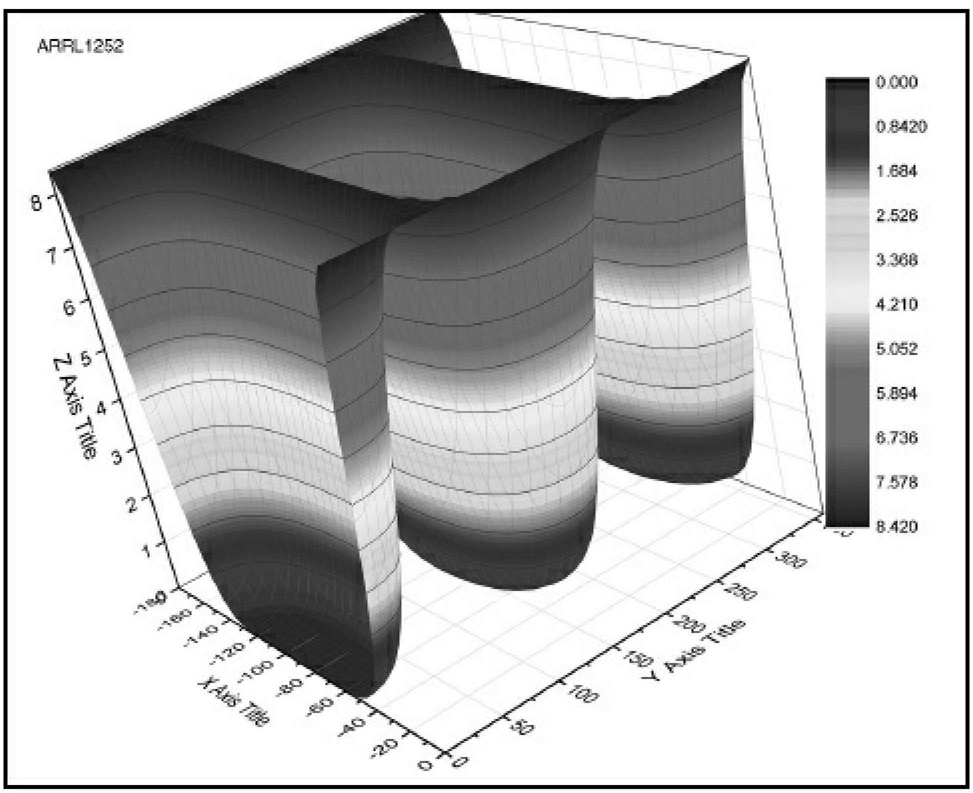


Figure 15.5 —Generated with *OriginLab 9.1*, this plot shows the relative field strength versus azimuth and elevation of an inverted V antenna over real ground.

Wish List

While ham radio propagation software is widely available, there are still a few computer tools that we *wish* were out there. If you're an enterprising ham or code poet, you might look into developing such a product. Here are just a few suggestions.

1) *More X and O propagation tools*. We mentioned above that *Proplab Pro* is the only current offering. It would be nice to have an open source version with the same functionality, (*Proplab Amateur*?) not *just* because hams are cheap, but because they always like to tweak things to make them better. It has been aptly said that open source software is "the new homebrew."

2) *Real time adaptive control*. This is software that would automatically adjust beam headings, frequency, polarization, power levels, and other factors, based on feedback from the distant receiving station. Shortwave broadcasters have used this technology for a long time to compensate for the vagaries of HF propagation. This is an area where hams can make a real contribution to the state of the radio art.

3) *"Intelligent" diversity reception methods*. We describe earlier how diversity reception has been around for a long time, but there are digital and mixed-mode methods available now that can greatly improve on these tried and true methods.

Keeping Up With Kepler and Friends

Unlike most aspects of radio propagation we have discussed, satellite propagation is pretty much like clockwork. While the ionosphere and lower regions of the atmosphere can affect satellite propagation to some degree, these effects are relatively minor. The biggest challenge for most radio amateurs when it comes to satellite communications is locating the satellite in the first place. Because of the ready availability of useful satellite tracking software, even this is not the challenge it used to be.

In addition, with the plethora of low Earth orbiting amateur satellites ("LEOs"), and general improvement in amateur station hardware performance, particularly on the receiving end, satellite communications is just downright simple. Gone are the days (with some infrequent exceptions) of having to steer massive dishes or complex arrays in the hope of achieving a fleeting glimmer of an amateur satellite signal. Armed with nothing more than a handheld VHF/UHF transceiver, and a handheld Yagi antenna, countless hams are making routine satellite communications.

Nevertheless, understanding some of the classic problems of satellite communications will increase your success and enjoyment of existing satellites and modes, as well as any future birds and modes that may come into existence.

Orbital Mechanics in a Nutshell

Figuring out satellite orbits isn't rocket science. Oh, wait. Yes, it is, come to think of it! Fortunately, it's a great deal easier to take advantage of satellites in orbit around the Earth than to put them up there in the first place.

By definition, any device in orbit around another is a satellite. The Earth is a satellite of the Sun, and the Moon is a satellite of the Earth. Commercial and amateur satellites (along with a lot of other space junk) are in orbit around the Earth, as well.

Once a satellite is in orbit around the Earth, there is only one force in play keeping it there, and that is gravity. There is also one force that wants to prevent it from staying up there, and that's friction.

Now, some will protest that it is *centrifugal force* that's keeping the satellite up there. But actually, there is no such thing as centrifugal force. Centrifugal force is merely the apparent *reaction* to *centripetal force*, which is the force that gravity exerts on the craft. Once the craft is in orbit, and the booster engine has stopped boosting, there is no change in the *kinetic energy* of the orbiting satellite. The centripetal force is always *at right angles* to the movement of the satellite, so it neither adds nor subtracts any kinetic energy from the satellite. In the absence of any friction (which would only occur if space were a perfect vacuum), any satellite would continue to orbit for eternity.

With that short physics lesson out of the way, let's talk about things that actually matter to the amateur satellite operator. Orbiting satellites of any kind always orbit in an *ellipse.* In some cases, the ellipse will be highly *eccentric*, as in the case of many polar orbiting satellites. In some cases, the ellipse will be nearly non-eccentric, or circular, as in the case of a *geostationary* satellite. But even commercial geostationary satellites have *some* eccentricity, which can contribute to a small amount of

Doppler shift, which can have a significant effect on telecommunications links and such.

Depending on the type of orbit (degree of eccentricity) and the *plane* of the orbit relative to the Earth, the orbital mechanics can range from very simple to rather complex. The simplest orbit to figure out is the geostationary orbit, where a satellite maintains a relative fixed position over a point on the Earth's equator. Obviously, a geostationary path can *only* be in an equatorial plane. Unfortunately, the geostationary orbit is one of the most expensive orbits to obtain, since the bird has to be some 22,236 miles high (very approximately the circumference of the Earth, by coincidence only). It is unlikely anyone is going to launch a geostationary amateur satellite any time soon.

Most amateur satellites are in much lower orbits than geostationary satellites, and can be in a plane far removed from the equatorial plane. Interestingly enough, the plane of any orbit will remain unchanged once the satellite is launched. If a satellite is in a polar orbit, for instance, the plane of the orbit will remain fixed (relative to space), while the Earth rotates around underneath it. This can, obviously enough, require some interesting mathematics to figure out where the satellite is going to be at any time relative to your location. And this is what *Keplerian orbital elements* or "Keps" are all about.

Keplerians are generally referenced to an *epoch,* which is a specific point in time at which you want to know what's happening. Since orbital mechanics don't really care about man-made time zones (or even solar time of day), you will first have to recalibrate your time to the epoch.

For a very detailed description of Keplerian elements, you can go to the AMSAT site: **www.amsat.org/amsat/keps/kepmodel.html**. This is highly recommended reading for the dedicated amateur satellite operator. But to get you up and running, the following abbreviated descriptions will give you a good basic understanding of how to find a satellite. These terms are fairly standardized on any satellite software you're likely to encounter. However, there are some optional elements that may or may not be required (or available) on some. Keep in mind the Earth is round, so you may have to alter your flat Earth thinking to make sense of all of these, but it's not hard at all with some practice!

Epoch: As described above, this is simply the time that the data is referred to, generally designated as T0. (That's T zero, not T "O").

Orbital Inclination (I0): Likewise, this is "I zero," not I "O"). This element is the departure of the satellite orbit from an equatorial orbit. Equatorial would be 0 and a polar orbit would be 90. However, the number you input into your software can range from 0 to 180. How can this be? Well, it depends on which way the orbit is "leaning." Theoretically, you could assign a + for a "right lean" and a – for a "left lean," but convention dictates a scale of 0 to 180 degrees.

Right Ascension of Ascending Node (O0): While the angle of inclination tells us how far the orbital plane is leaning, it doesn't tell us which *axis* it's leaning around. Imagine driving a broomstick through a hula hoop at two diametrically opposed points. If the broomstick is horizontal, the Right Ascension is the "azimuth" of the broomstick. Now, even if we orient the broomstick based on a 360 degree circle, there will still be some ambiguity of which direction the satellite is in orbit…in other words, which way around the hula hoop it's traveling. Again by convention, we define Right Ascension as the south to north direction of the satellite, based on your position on the Earth. Obviously the descending node is going to be on the other side of the Earth going from north to south. Now, keep in mind the Earth is also rotating underneath this orbit, so depending on the time of day at your position on the Earth, a Right Ascension can be a Descension. To avoid possible ambiguity, this plane of ascension is not defined relative to the spinning Earth, but to a fixed point in space.

Argument of Perigee (W0): Since all orbits are elliptical, we need to define the direction the ellipse is "tilted." It's handy to know that the center of the Earth will always be at *one* of the focal points of an ellipse. This means we only need to define a "non-Earth-centered" point to define which way the axis of the ellipses is tilted. We *could* use the second focus of the ellipse, but this really has no practical significance. Instead we define the point of perigee, which is where the satellite is closest to the Earth. The complementary point is the apogee, where the satellite is farthest from the Earth. If we define the perigee, we can determine which way the ellipse is "leaning." This lean angle is, by convention, between 0 and 360 degrees. (By now, you are probably wondering why there isn't much consistency in these definitions. It could be said that they're *uniformly* inconsistent. What we mean is, as odd as some of these definitions are, everyone *now* uses the same oddities!)

Eccentricity (E0): This figure tells you how far the orbit departs from a circle. A value of 0 is a perfect circle, while a value of 1 is a line. This definition is at least perfectly consistent with the term you learned in high school geometry. (Whew!) This figure is simply the ratio of the minor axis to the major axis. A value between 0 and 1 is a normal ellipse.

Mean Motion (N0): So far, we've pretty well defined the orientation and shape of the orbit. But we haven't said anything about how *big* the orbit is. The orbital size has the most significant effect on the *period*, or how often the bird will be visible in a given period of time. We know from Kepler's Third Law that there is an inverse relationship between orbital size and period. The simplest way to express this is by the *period*; that is the time it takes the satellite to return to an exact location in space. This is a simple figure given in revolutions per day.

Mean Anomaly (M0): While *anomaly* is normally associated with something *abnormal,* in this sense it simply means the difference between where the satellite is and some reference point. In this case, our reference point is the Epoch, or the starting time of whatever orbit we're looking at. The location of the satellite at any point of time is "phase shifted" from our epoch by some value. This is defined by a 360 degree circle, with the center of the circle being at the center of the Earth.

Drag (N1): This is an optional figure. Notice the "1" designation as opposed to a "0" or fundamental designation. There isn't much you can do about this; it simply tells you how much each orbit is going to change from the previous one. For the typical amateur, this is pretty meaningless, while for the actual owner of the satellite, this means a lot more. Your regularly updated Keps incorporate this figure in their calculations.

There are a few other parameters that take into account the bumpiness of the Earth's gravitational field and such, but you will probably not be too concerned with this, unless, again, you're a launch mechanic.

The ARRL and other sources provide regular bulletins of Keplerian elements. Most amateurs use one of the satellite tracking software packages to handle the calculations and even control rotators. See **Figures 16.1** and **Figure 16.2** for examples.

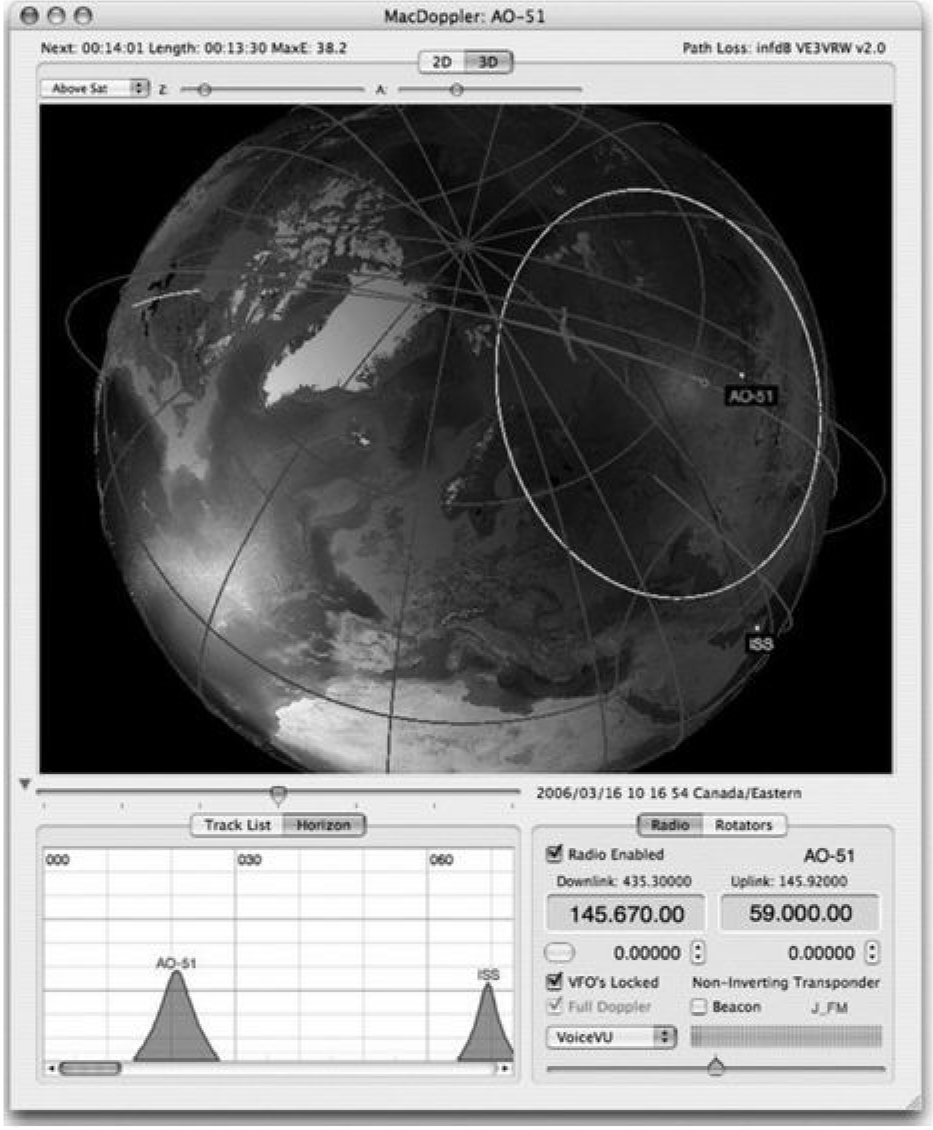


Figure 16.1 — *MacDoppler* is a professional grade satellite tracking program that will interface directly with a number of transceivers to adjust the tuning and Doppler shift, so as to make tracking a satellite a snap.

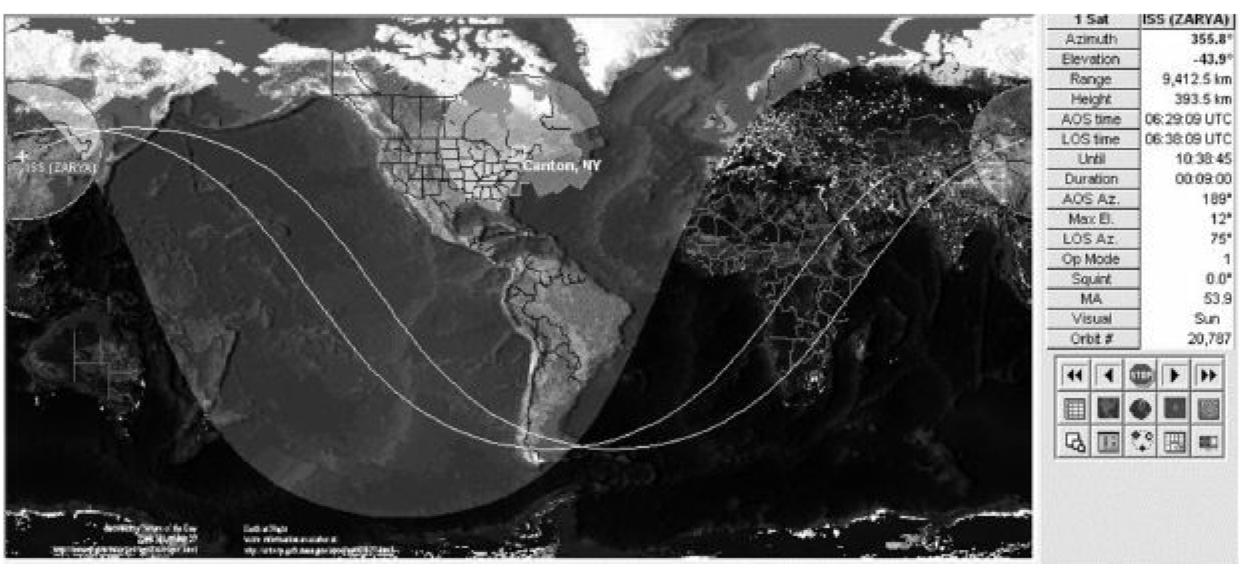


Figure 16.2 — *Nova* for *Windows* is an older tracking application that is still very popular. It has automatic az-el antenna steering capacity if you have the correct interfacing hardware. Autotracking hardware is actually quite simple to set up if you already have an az-el steering system in place.

A Word of Note

In a lot of satellite literature, you'll hear it said that as an orbit "decays" the satellite will "speed up." By "speeding up" it is meant that the *period* will shorten, that is its Mean Motion increases. This does *not* mean that the velocity of the satellite is increasing in its direction of travel! Friction (due to the atmosphere, as thin as it is up there) burns up kinetic energy, causing the velocity of the satellite to decrease. This concept suffers from nearly universally lame usage of terminology. You cannot violate conservation of energy or Kepler's Laws!

Some More Polarizing Words

Every satellite user will eventually encounter the ambiguity of *polarization* terminology. Polarization is typically defined as *vertical* or *horizontal*. "Vertical or horizontal compared to *what?"* you rightly ask. It's easy to define the polarization of a radio signal traveling parallel to the surface of the Earth. If the electric field is perpendicular to the ground, it is vertically polarized; if it's parallel to the ground, it's horizontally polarized. No problem.

But what if the signal is traveling *perpendicular* to the Earth's surface, which describes most satellite signals? In this case, you can only define polarization in terms of north-south, or east-west. Even this isn't entirely satisfactory, but it at least gives us a starting point. Assuming you are on the Earth, standing in the orbital plane of the satellite, a north south polarized signal will have its electric field oriented in an actual north-south direction. Likewise an east-west polarized signal will have its electric field oriented in a genuine east-west direction.

However, if you are *not* in the orbital plane, things get a lot more interesting. First of all, as we mentioned in our chapter on polarization, the *polarization sensitivity* of any antenna decreases as you move off of "boresight." Secondly, the optimum polarization orientation of your receiving antenna (assuming the satellite is the transmitter) will change, depending on how far off boresight you are.

In addition, if you are *not* in the orbital plane, the departure from N-S or E-W orientation will be dependent on your *latitude* as well! All this is further complicated by the fact that even *commercial* satellites refer to the meaningless horizontal and vertical polarization terminology. What's a ham to do!

Fortunately, the situation is not as desperate as you might think. As in many other cases, it's the very *imprecise* nature of most practical antennas that makes things work "better than they should." In the case of satellite reception, your chances of being perfectly cross-polarized with a bird are next to nil. Once you have *some* kind of signal, you can always rotate your polarization to get a *better* signal, without even having to worry about definitions. Just be aware that once you have things tweaked for best operation, your antenna polarization will not likely be identifiable as either horizontal, vertical, N-S *or* E-W!

Circular Reasoning

For the above reasons, and others as well, many hams use circular polarization for satellite communications. A circularly polarized ground station antenna doesn't care what the polarization of the satellite's signal is. One of the most common uses of circular polarization is when you don't have the desire or capability of *tracking* a satellite, and yet still want to get a reasonable signal throughout the duration of its pass. As we've shown the relative polarization of a satellite's signal can change dramatically over its path. A circularly polarized, broad pattern antenna such as the bifilar helix will be relatively unaffected by the shift in polarization as the satellite progresses in its orbit.

Faraday Rotation

Another reason to use circular polarization is to minimize the effects of *Faraday rotation.* Faraday rotation is the rotation of the plane of a linearly polarized wave as it travels through the ionosphere. This is very *different* from X and O propagation, in that a Faraday rotated plane wave is still a plane wave. X and O waves as described earlier are circularly polarized. There is often great confusion over this matter. However, these two phenomena have entirely different causes, and different manifestations.

Faraday rotation is seldom appreciable at UHF frequencies, but increases as the square of the wavelength. It can be quite noticeable on 2 meters while nearly undetectable on 70 cm. The degree of Faraday rotation is also directly related to the Earth's magnetic field, as well as the number of free electrons available. As such, Faraday rotation can be an extremely sensitive indicator of sporadic E, meteor burst, and a number of upper tropospheric phenomena. The real difficulty in using Faraday rotation for measuring such phenomena is finding a stable source of plane polarized waves to start with. Interestingly enough, moonbounce signals are one of the most promising sources of stably polarized signals, although they tend to be a bit unstable in other regards.

But, While We're on the Moon Anyway

Moonbounce communications until relatively recently were near the extremes of Amateur Radio ambition. In the early days, it took massive antennas, exotic receivers (sometimes cryogenically cooled) and absolute maximum legal RF transmitter power. Astonishing technological improvements, especially on the receiving end, largely due to extremely low noise devices such as GaAsFET preamplifiers and the like, have changed the game entirely. In addition, very low noise, weak signal digital processing protocols, such as those found in *WSJT* software, have made moonbounce nearly as accessible as amateur satellites.

In Chapter 3, we described the depressing radar equation and how it applies to reflections from relatively nearby objects. When considering moonbounce, we need to extrapolate the depressing radar equation even further. The typical path loss for an EME (Earth-Moon-Earth path is on the order of 250 dB. That's an awful big amount of loss!

Now, here's something quite interesting. A polished metal sphere, regardless of diameter, has a radar cross section of essentially a pin point. Geometrically, this is very simple to visualize. If you shine a laser pointer at a 3 foot diameter polished steel sphere in a dark room, the only time you'll see a reflection is when the beam is *precisely* perpendicular to the surface of the sphere! Any reflection from anywhere but dead center is scattered somewhere else, and you'll never see it. So the effective "size" of the sphere is essentially zero.

Now, if this is the case, how do we get *anything* back from the Moon? As it turns out, it's the very *roughness* of the Moon's surface that saves the day! No matter where a radio signal hits the Moon, the rough surface is going to reflect *some* of it back in the same direction it came from. If we have millions and billions of little crevices and bumps, even if they scatter most of the signal somewhere else, there will be a certain quantity (statistically speaking) that bounce back toward the source *in phase*. In effect, the Moon, as far as radio goes, looks *nearly* the same size as a disk of its diameter. Now, the downside of this is that the Moon is not too stationary. It tends to "rock" in its orbit a bit, which means the relative distance of all the countless reflecting surfaces change slightly, causing the relative phasing of the reflected waves to shift. This is called *libation fading*. And, as annoying as it is, it's a small price to pay for *no* detectable reflection!

Now, here's just one more little kink in the works. The *subtended angle* of the Moon is about half a degree. What we mean is that if you stood in your back yard and took a protractor and measured the angle difference between the "east side" and the "west side" of the Moon, you'd have about a half degree. Now, a really good Moonbounce antenna may have a half-power beamwidth of about a degree. This would be a world class installation with perhaps a 30 foot dish. Most moonbouncers are probably using antennas with between 1 and 2 degrees of beamwidth. This means the Moon only intercepts a fraction of the RF aimed in its direction. So, this makes the Moon appear a little "smaller" than it would with a perfectly matched beam pattern. Again, this "slop" actually makes it a little easier to find the Moon in the first place, but it does contribute materially to the signal loss.

Despite all of this, hams have been doing moonbounce for about half a century.

The article, "EME on a Budget: Moonbounce for the Rest of Us" by Paul Bock, K4MSG, gives some valuable insights into this. See **www.k4lrg.org/Projects/K4MSG_EME/.**

Really Out There

No chapter on space communications could be considered complete without making at least passing mention of deep space radio. Radio amateurs are *still* receiving radio signals from Voyager 2, the deep space probe launched in 1977, and now traveling through the *magnetosheath* region of the outer regions of our solar system. This is astonishing for even the most jaded science geek. While it speaks volumes about the technology, it also is a testament to the amazing logarithmic nature of electromagnetic radiation itself. Theoretically, it's possible to receive a radio signal forever; its strength just keeps dropping off as a square of the distance for infinity. Of course there are *practical* limits to this, but it still never ceases to astound, how well radio works!

It is probably not outside the realm of possibility that fully amateur "Mars bounce" should occur. This obviously would require some serious signal processing — not to mention some really long integration times — but the government has been talking to Mars craft for a long time. It's really only a matter of scale.

Practical Pointers

Back on Earth, let us tie up some loose ends with regard to Kepler. As we explained, Keplerian elements will tell you where you need to point your antenna to hit your satellite of interest. It's one thing to know where to point your antenna. It may be another thing entirely to actually be able to do it.

When working with a very large array (especially one you built yourself), it's easy to assume that the main lobe of the antenna is physically where the antenna is pointing. For a single high gain Yagi, this is probably reasonably true. The fact that countless hams are talking to birds with hand held Yagis testifies to this.

Some more exotic satellites (such as the Moon!) may need a little more elaborate means of antenna pointing. For instance, a large array of phased Yagis (a rather typical moonbounce setup) requires very close matching of the phasing harnesses and such. It doesn't take much error to throw the main lobe several degrees off of where you *think* the thing should be pointing. (This applies to any phased array, actually.) Aiming by "eyeball" you can still miss the Moon entirely, unless the array is extremely well constructed.

One way you can verify the alignment of your antenna is by using the Sun as a noise source. If your receiver is adequate for receiving moonbounce, it is certainly capable of receiving solar noise. Peak the antenna for maximum solar noise, and then use that to boresight the Sun optically (using adequate eye protection, of course!).

Traditionally, a sighting tube was used to align an antenna with the Moon. However, inexpensive and sensitive tower-mounted CCD cameras make the whole process that much easier. You can steer the array from the comfort of your warm shack, rather than standing out in the yard at the foot of the tower, staring up through the bore sight, screaming at your assistant (if you can find one), poised at the rotator controls.

Where to from Here?

The predictability and reliability of satellite communications can give the impression that there really isn't much to learn about radio propagation, at least compared to HF radio, for instance. This assumption can be deceptive, however. Subtle effects such as Faraday rotation, scintillation, and even fading, can reveal some fascinating things about our universe. It is precisely because satellite communications *are* so predictable that they form a great reference point for observing the new and unique. Insatiable curiosity and a scientific mindset will keep amateur satellite activity exciting for a long time to come.

Your Friend the S Meter

Since about 1940 or so, radio amateurs have declared the standard signal strength meter, or *S meter*, to be obsolete. And yet, there hasn't been a commercially made HF transceiver built since at least that time that hasn't included an S meter with essentially identical characteristics as that of its predecessor of 70 years ago (**Figure 17.1** and **Figure 17.2**).

The fact that it's been around for so long is a pretty good indication that the device was well conceived in the first place. Nobody's come out with anything that's demonstrably *better*, when it comes to evaluating propagation conditions on a large scale. Let us look at the S meter in detail and show why it is still unswervingly *useful* in light of known ionospheric propagation characteristics.

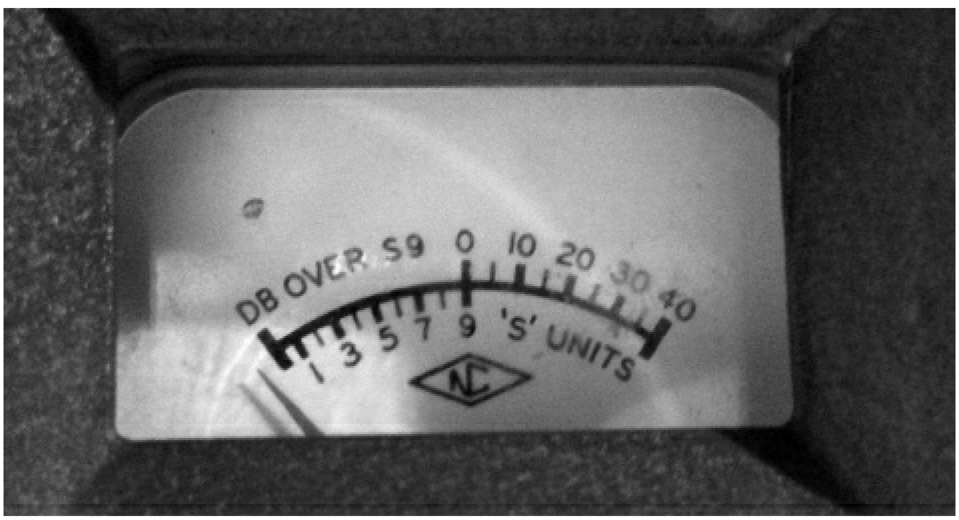


Figure 17.1 — The S meter of a classic National NC-109 receiver, with a full scale deflection of 40 dB over S9.

Figure 17.2 — A Heathkit HR-10 S meter, with a full scale deflection of 60 dB over S-9. Vintage receivers can have quite accurate S meters, providing they are calibrated at the frequency of operation. Modern receivers tend to have much more uniform response throughout the HF spectrum.

An Old Standard

The standard issue S meter incorporated on most amateur grade HF receivers and transceivers has fairly uniform characteristics. The S meter has a linear scale beginning at S-1 on the bottom end, and extending up to S-9 at about 2⁄3 scale. Above S-9, the meter has a logarithmic scale with divisions in terms of decibels above S-9, generally in increments of 10 dB. The greatest variation in S meters is typically where the meter tops out. Older communications receivers such as the semi-venerable Hallicrafters SX-100 had a scale that went up to +80 dB. Most modern transceivers are a bit more pessimistic at the top end, often going no higher than +50 dB or +60 dB.

With a well calibrated S meter, each S unit below S-9 represents a change of 6 dB, or a halving of antenna voltage. One interesting feature of this arrangement is that below S-9, the meter is indicative of *voltage,* while above S-9, it is indicative of *power.*

Power Play

It should be noted that the S meter had become largely standardized well before the adoption of coaxial transmission lines by the average ham. The standard input of most HF receivers was a pair of 300 Ω screw terminals, either balanced or unbalanced.

Nowadays, of course, the nominal input impedance of a communications receiver is 50 Ω, not because that has any particular advantage for *receiving*, other than it allows things to be standardized across the board for 50 Ω coaxial cable. But let's step back a few years and look at what, if anything, was magical about 50 µV being S-9. We can actually thank the phone company for this.

Now Hear This

When Bell Telephone first started standardizing things, they determined that a good listening level for a standard telephone headset occurred with an input power of –13 dBm (dBm is decibels relative to a milliwatt; 1 mW is 0 dBm). In fact –13 dBm is *still* standard operating level for twisted pair analog land lines, where such relics still exist.

By the time the S meter rolled around, the superheterodyne receiver had become fairly standard. (You won't find many S meters on vintage regenerative receivers, for instance, though there might have been a few in existence). A typical superhet receiver had an overall gain, if you included the RF, IF, and audio gain together, of about 60 dB, or a 1,000,000:1 power gain. This could be achieved with five or six vacuum tubes, a typical run-of-the-mill communications receiver lineup.

Let's work backward from the audio output and come up with some numbers. (By the way, –13 dBm comes out to about 0.00005 W). Recall that the overall power gain to get us up to 0.00005 W is 1,000,000:1. So we can divide 0.00005 by 1,000,000, and we come up with 0.00005/1,000,000 = 0.00000000005 W. Or 50 *picowatts*. Now, remember, this is for an S-9 signal. Now, just as a sanity check, let's work this out from the antenna end.

A modern receiver has a nominal 50 Ω antenna input impedance. What input *power* does 50 *microvolts* across a 50 Ω load represent? Well, using E squared over R (E^2/R, the standard power formula with a known voltage and resistance), we have (0.00005 × 0.00005)/50. Lo and behold, when it all comes out in the wash we have 50 picowatts. That seems to match up.

Now remember, a watt is a watt. 50 picowatts is 50 picowatts. An S-9 signal is 50 picowatts. The fact that ancient receivers didn't have 50 Ω input impedance is irrelevant. An S-9 signal was the input power level that, with typical gain, could give you a typical "listenable" signal in a Bell standard telephone earpiece, regardless of antenna input impedance. It just turns out that with a modern receiver having an input impedance of 50 Ω, that 50 picowatts corresponds to 50 µV. And so, because most modern receivers have a nominal input impedance of 50 Ω, we, by convention, can calibrate our S meters in microvolts instead of picowatts.

If you want to drive a big speaker, you'll have to add 30 dB or so of audio power gain. Incidentally, a receiver with 1 W of audio power output is generally enough to blister the paint off the walls of a typical ham shack. We won't include this gain in our following discussions, but can refer to this as "outboard" gain.

And Some AGC

Actually, by the time the S meter came about, most receivers had a lot more than 60 dB of gain available, but they weren't generally run "flat out." Almost from the beginning, the S meter was AGC-derived, that is, it was driven from an automatic gain control dc bus, which was used to automatically reduce the gain, in order to keep signal levels somewhat uniform. In reality, this means that there's a considerable amount of gain reduction in play for an S-9 signal in any typical receiver.

An interesting byproduct of using AGC derived S meter readings is that there's a built-in logarithmic function in the process. This log function is true of any negative feedback amplifier, where the feedback itself is variable. The S meter voltage itself is a logarithmic function of the antenna voltage. This allows the S meter to cover a much wider *dynamic range* than would be possible if the S meter were simply a sample of headphone voltage, for instance. While the log function is in place any time there is AGC action, it's a lot more noticeable at the "loud" end of the scale, which means that below S-9, the scale can be somewhat linear, while above S-9, it could be highly compressed.

None of this feedback theory has changed in the last 70 years, by the way; we just use different technology to do the job nowadays! There were some really smart people who worked out this S meter business, and there hasn't been anyone sufficiently smarter since then (that we know of) to justify abandoning it completely.

The Rise and Fall of the Signal Report

It really is a shame, in light of all the thought that went into the development of the S meter, that for most hams it amounts to little more than a hood ornament. While it's true that for typical amateur communications the *intelligibility* of a signal relies more on *signal-to-noise* ratio than actual signal strength, the S meter is all important for doing anything meaningful with propagation investigation. (At this juncture we probably need to acknowledge that the universal S-9 "contest report," as annoying as it is, probably is as good as it needs to be. You either copy the station or you don't).

But that's not what the issue is in this discussion…or in this entire book. Any kind of meaningful propagation study requires that you look at actual *signal strength.* And for probably 95% of our readers, the S meter is (or can be, with some help) the most accurate indicator of signal strength to be found in the amateur station. A small number of radio amateurs *do* have actual, calibrated, absolute field intensity meters available, but these are expensive and rare devices. Fortunately, for the bulk of what we do in propagation studies, a sensitive, stable, and repeatable *relative* signal strength indicator is all we need, and the S meter fills the bill in style…with a few caveats.

Below S-9: Where the Action Is

While most hams like the thought of their signals slamming the distant station's S meter against the ceiling (not necessarily an ignoble goal), when it comes to actually measuring things, it's what's below S-9 that counts. First of all, between S-0 (if your meter actually has such a marking) and S-9, the S meter is *linear.* And, furthermore, it's directly related to antenna voltage. In addition, each S unit represents a halving of antenna voltage (6 dB). This is an extremely convenient scaling, by the way, for a number of reasons.

We know that the field intensity (voltage) of an electromagnetic wave decreases by the function of 1/R, where R is the radial distance from the source. (Remember that at any significant distance from a transmitter, the wavefront of a radio wave becomes spherical.) This means as you double R, the field strength drops in half, or by 6 dB, or 1 S unit. Remember the formula for decibels when *voltage* is concerned is:

$$dB = 20 \log(V1/V2)$$

When using dB to express voltage ratios, also recall that both V1 and V2 have to be across the same *impedance.*

You will also recall that the available *power* attenuates as the inverse *square* of the distance, so each S unit represents a power ratio of 4:1. However, voltage is much easier to measure directly than power, so for our purposes voltage ratios are perfectly fine to use. We just have to be sure that comparative measurements are made using the same impedance, in this case the input impedance of our receiver.

Real Receivers

We need to recognize that the 50 Ω input impedance of a receiver is a *nominal* value. In the case of most modern, broadbanded, solid state HF receivers, the *actual* input impedance of a receiver is pretty close to that. You can easily verify this if you routinely use an antenna tuner with a transceiver. If you tune your antenna tuner for the best received signal of a distant station, it will also be a good match to your transmitter. As they say, your mileage may vary. However, at any given frequency, the input impedance of a receiver will be the same regardless of input level — up to the point where the front end saturates. For measurement purposes, we will be operating well below this point.

Older vacuum tube receivers may have much wider variations in antenna input impedance, especially from band to band. This is one of the main reasons the S meter achieved a bad rap once hams entertained the notion of obtaining actual absolute field strength measurements. You couldn't really derive actual antenna input power without knowing what the input impedance was. Hold that thought for a moment.

Infinite Impedance

One of the ways you can accurately measure field intensity without accurately knowing the load impedance of the receiver is to use a receiver with *infinite* input impedance. In reality, this is the ideal instrument. The so-called "active antenna" closely approximates this ideal. Basically we want to create a *voltage probe* that samples the electric field intensity of the incoming radio wave without disrupting that radio wave. This is the same principle of using a high impedance voltmeter to measure high impedance electronic circuits. In an active antenna, there is no effort made to achieve impedance matching; its purpose is to merely "sample" the radio wave.

This is decidedly different from your impedance matched receiver, where the goal is to extract as much power as possible from the incoming radio wave, in order to maximize the sensitivity of the receiving system.

Let us return now to our vintage vacuum tube receiver with the widely varying input impedance. Is it still possible to make valid field intensity measurements with such a beast? A qualified "yes" would be the answer. We can certainly make valid *relative* measurements. In most propagation studies, we only need to know the direction (sign) and degree of *change* of signal strength. If we can consistently demonstrate that a 2:1 change of field intensity results in 1 S unit of change on our S meter (at one given frequency) we can have a powerful instrument.

One of the most reliable means of independently verifying the linearity and scaling of an S meter is by using external calibrated RF *attenuators.* A number of excellent switchable (ladder) attenuator projects have appeared in *QST* magazine and the *ARRL Antenna Book* over the years. Any dedicated experimenter should have a precision attenuator in the shack. It also helps to have a well-calibrated RF signal generator around, but this isn't absolutely necessary if you have a steady radio signal available, such as from WWV.

What you want to do is simply see if switching in 6 dB of attenuation results in a change of 1 S unit, at your frequency of interest. Ideally, you want to test this out for every increment between S-1 and S-9. If your S meter does not track your attenuator settings, you have several options. One of these is to build up a *calibration chart*. This is simply a spreadsheet or graph that shows what S meter reading corresponds to which actual amount of attenuation. It can be a bit tedious do develop a calibration chart, but this sort of thing was standard practice in a lot of ham shacks (and professional labs) long before computers were around.

Another alternative is to use *only* your attenuators. The way to do this is to always adjust your attenuators to achieve some S meter reading, say S-6. As the signal under test changes, you simply adjust your attenuators to retain an S-6 reading. The amount of attenuation you add or subtract will be equal to the actual variation in signal strength. This is probably the most reliable method, since the *linearity* of the S meter circuitry becomes irrelevant. The downside is that it can be a bit inconvenient from an operational standpoint.

Swampland

We mentioned above that one way to make the receiver input impedance irrelevant is to use an active antenna with *infinite* impedance. Another method is to *pad down* or "swamp" the front end of the receiver with lots of fixed attenuation. This method greatly levels out the input impedance of the receiver, but at the cost of greatly reduced sensitivity. Again, for measurement purposes, we aren't trying to achieve maximum sensitivity; we just need enough signal to make valid measurements.

By the nearly magical means of swamping, even the most decrepit boat anchor receiver can exhibit an astonishingly flat front end. The liberal use of padding and swamping is standard practice in the microwave industry, by the way. Most microwave directional couplers and filters require the generous use of padding to function properly. Efficiency and precision are pretty much inversely proportional in the microwave lab!

It's All Relative

From a purely practical standpoint, as radio amateurs, we want to know if Action A results in Result B getting better or Result B getting worse. We want to know if doing a certain tweak improves or degrades some aspect of performance, and by how much.

On the other hand, while *absolute* measurements may be interesting from a radio science perspective, we can often get by with little reference to absolute measurements. How often do we tell a DX station, "Your signal is 50 µV per meter here, OM"? But informing the other station that the signal from a new phased array is 3 S units above a Yagi is something really meaningful…especially if you know your S units are real!

Loads of Modes

Often when considering ionospheric propagation, it's tempting to just assume that "RF is RF" The actual modulation applied to a radio frequency carrier would, on the surface, seem to have very little to do with how the radio signal travels from Point A to Point B. This is largely true, but as we discovered in the chapter on diversity methods, there can be rather sudden and profound differences in propagation over the space of just a few kilohertz…sometimes less.

This phenomenon really came to light during the early years of HF packet radio. Unlike many instances in Amateur Radio when things seem to work a lot better than they should, HF packet radio almost *never* seemed to work as well as it should have. Even when signals were S-9++ going both directions, HF packet was plagued by countless retries, slow data rates, and all the other maladies that could possibly afflict digital radio. This is not to say that HF packet doesn't still exist. There are still a few lone chirps to be found on the HF bands. But for the most part it suffered a long, slow gasping death, as nearly every other possible HF mode seemed to work better.

For a while, there weren't any satisfying explanations for why HF packet didn't work as advertised. The error correcting protocol was robust. The nominal data rate was not excessive. It was a well-conceived protocol.

As it turns out, not many in amateur circles recognized just how rapid and selective ionospheric fading could be. Some well-seasoned RTTY operators had some hints, but they never really pushed the limits of data rates as packet did. (It may be hard for anyone raised on the Internet to consider 300 baud as pushing any kind of limit, but such was the case!)

The ionosphere can undergo *tremendous* amounts of Doppler shift — enough to significantly skew ionospherically reflected signals in both frequency *and* time delay. After a lot of investigation (not only by radio amateurs, but also by professional radio scientists who were also dabbling in packet techniques) it was determined that the *real* culprit was *intersymbol interference* caused by highly selective time-of-flight errors. There was enough of a difference in reflection height between mark and space frequencies that they "slopped over" each other more often than not. No increase in signal strength or signal-to-noise ratio could help remedy this one iota. And, as it turned out, this effect would *just* come into play at around 300 baud; it was seldom noticeable at typical 45 baud RTTY rates. Of course, the effect was progressively worse at higher data rates.

Clover

Clover, a very expensive (at the time) protocol, was the first HF digital protocol to actually recognize this intersymbol problem. It was first made readily available by HAL Communications, a company famous for bringing high quality RTTY terminal units to radio amateurs. And it addressed the issue in some clever, innovative, and interactive ways. In fact, Clover wasn't a single protocol, but several. One of the most obvious was to use more, but more closely spaced, frequency shifts. But it also incorporated some frequency and time diversity methods. And it did some serious number crunching to figure out what was the best combination of "fixes" to apply at any one time.

Because of its relative expense, Clover fell off the radar in its early adolescence, but seems to be enjoying a bit of a revival, primarily because the necessary computer processing has gotten a lot cheaper than when Clover was first rolled out.

Another interesting feature of Clover is a very long sequence forward error correction (FRC) protocol, which slowed things down a bit, but was phenomenally reliable under the most abominable conditions. In FEC mode it far surpassed AMTOR, nearly matching the performance of AMTOR in ARQ mode.

AMTOR Notes

AMTOR deserves more than passing mention here. AMTOR was derived from the highly successful commercial maritime SITOR system. In fact, the systems are nearly identical. AMTOR's real power was its automatic repeat request (ARQ) connected error correction, which was the first amateur protocol to use this. Under ARQ the connected link would yield essentially 100% perfect copy…or nothing at all. AMTOR does have a few liabilities, besides the fact that you never hear it on the air any more. Like RTTY, AMTOR was limited to the standard 5 bit Baudot character set, which only included capital letters and numerals. The Internet generation would call this the "yelling mode." In addition, the rapid fire talk-and-listen ARQ mode required the transceivers at both ends to continually flop between transmit and receive a few times a second. If you had any TR or antenna relays it would drive you up the wall, and it didn't do your antenna relays and other peripherals much good either. If you had full break-in (QSK) it worked very nicely, but at the time there weren't a lot of full power solid state TR switches available.

Despite its shortcomings, AMTOR paved the way for PACTOR, "Packet AMTOR," which is pretty much the default standard protocol for automatic HF relaying, such as Winlink. PACTOR has the complete ASCII character set, as well as a lot of concepts borrowed from Clover, such as long sequence forward error correction, adaptive data rates, and multitone quadrature modulation (in the later versions). PACTOR has evolved a bit itself, and is now up to PACTOR 4, which is used primarily on commercial circuits.

And Some ALE

Automatic Link Establishment (ALE) is a framework developed by the military for automatically establishing HF radio links under a wide variety of conditions. It is not one specific protocol, but rather some broad concepts that *were* formulated into specific protocols by Winlink and others. An ALE system often uses an interactive propagation predicting program such as *ITS-HF* in order to optimize its frequency and data rates.

Ping Pong

In our earlier sections on channel probes, we described how important active feedback is in optimizing signal paths, something radio amateurs are just beginning to implement. While a channel probe (a beacon transmitter paired with a dedicated receiver) is useful for evaluating a known path, it can't tell you the best HF path of numerous options, or even tell you what those options are.

A refined ALE system, on the other hand, will scan the radio spectrum (at least the parts it is duly licensed for) and determine which frequency "works" for the required mission. This is called “pinging,” a term derived from SONAR, but also adapted wholesale by network administrators. Pinging was actually well developed by the government for meteor burst communications. Meteor burst paths are short and fleeting, but extremely reliable when they exist. Many remote scientific sites rely on meteor burst communications when no other infrastructure exists.

The concept is simple. A remote site will ping the sky (generally in the 5, 6, or 7 meter frequency range) for a radio path to the “home base.” If the home base receives a signal, it immediately fires back to the remote site, “We have a path!” upon which time the remote site sends a big data dump. Meteor burst is capable for extremely broadband data, but only at random intervals. Again, this concept is well developed for government and scientific purposes, but hasn’t been implemented much by radio amateurs, especially at HF, where the concept could be extremely useful.

Sound Cards Galore

One of the most popular sound card software packages is known as *fldigi*, which has 16 basic modes, not including the various data rates available for many of those modes. See **Figure 18.1**. A “more mature” all-mode package, known as *MultiPSK* (**Figure 18.2**), has nearly an equal number of modes, many of them essentially discontinued. One might ask why there are so many different digital modes; how many ways can one send text, anyway?

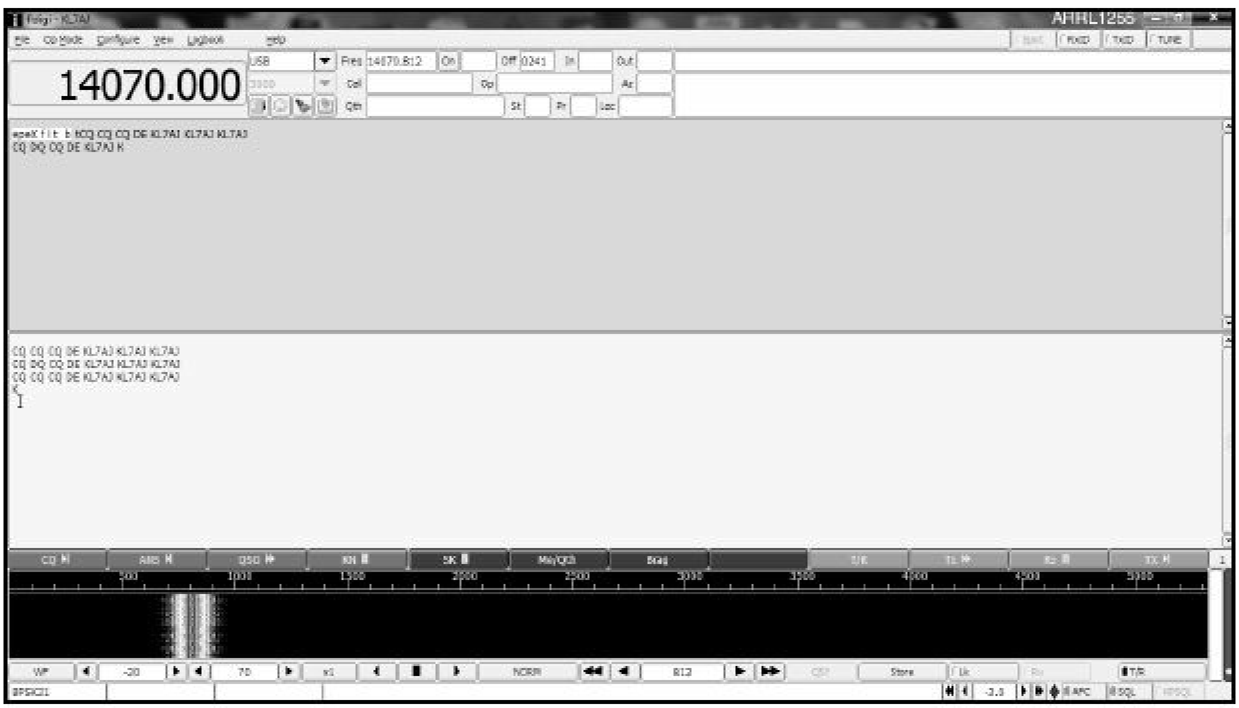


Figure 18.1 — *Fldigi* is probably the most popular all mode soundcard program. It’s extremely easy to set up, and it can even decode most signals just by putting a laptop’s microphone in front of a general coverage receiver’s speaker. It is amazingly immune to background noise. However, one should always “hard wire” their computer to their rig for good engineering practice.

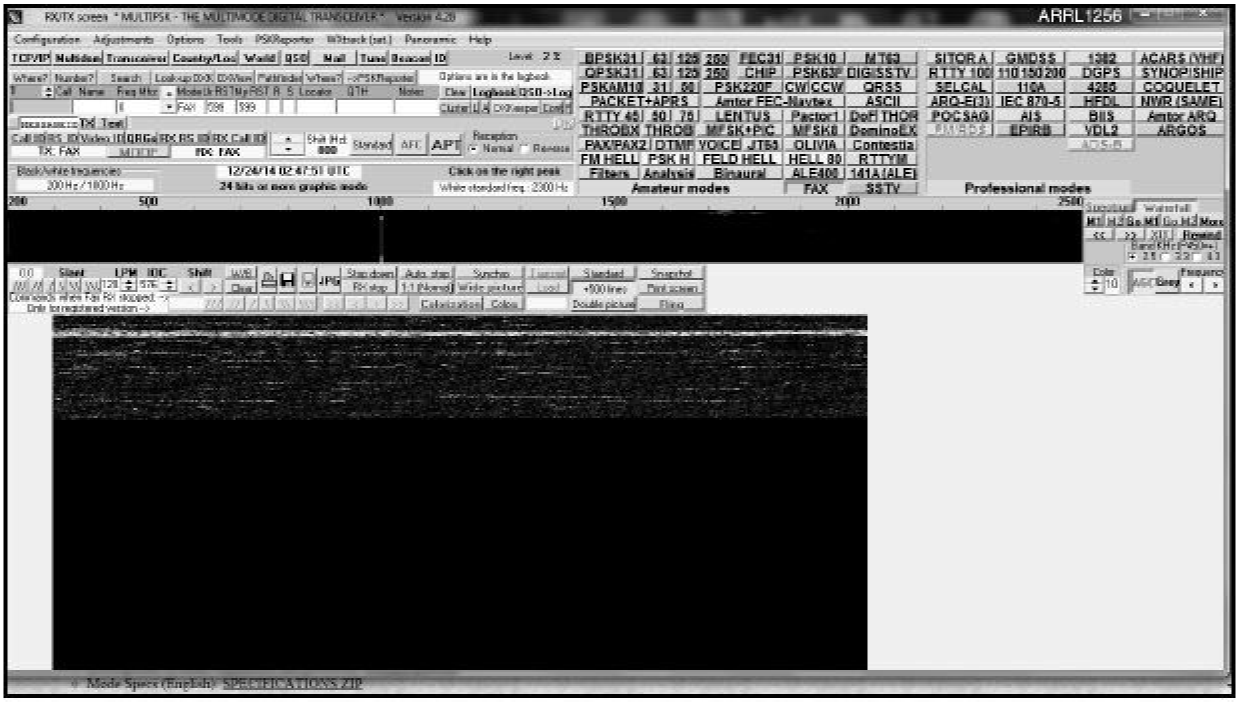


Figure 18.2 — *Multipsk* is an early all-mode program (with recent updates) that has a very busy GUI (graphical user interface). Nevertheless, it's a reliable, robust program that actually works on all the modes it offers. It also has a few modes most hams have never heard of. What more can you expect from a free program?

It is fairly safe to say that each mode was developed to compensate for, or take advantage of, a particular peculiarity of ionospheric propagation. The ever present drive for greater data throughput, compatible with available RF bandwidth, has instigated some truly unique, and sometimes truly odd modes. Some of them, like Olivia, even *sound* strange.

The peculiarities one must contend with are: fading, selective fading, Doppler shift, polarization shift, selective (differential) "time of flight" and various types of noise. These peculiarities may apply individually, or in combinations, and each digital mode deals with them with greater or lesser success. Since these software packages are absolutely free, it costs nothing to do a lot of experimenting, to find out what works best for you at any given time.

Error Correction

One of the interesting aspects of these different modes is how they handle error correction, and the overall philosophy related to this. As in any other engineering discipline, there are tradeoffs.

In a perfect world, you wouldn't need any error correction. If you could devise a modulation scheme that flawlessly compensated for all the known variations of ionospheric propagation (or other volatile propagation paths), you could get by without any error correction. Of course, in practice, the best you can do is make an educated guess, and hope that you've covered most of the bases. On the other hand, you could use a simpler modulation scheme (which almost dictates that it would be faster) that is much more subject to errors, but then use a sophisticated error detection and correction scheme, which takes processing power and time. Either way there is a cost, and that cost is *throughput* or ultimate data rate. Without simplifying the matter too much, we can narrow your choices down to just two:

- Saying it slowly and precisely once.
- Saying it fast and sloppy a whole lot of times.

As limited as these choices are, how to apply them, and in which proportion, can be a very complex engineering issue.

As an example, let's take the case of slow speed CW, which ostensibly has the best possible signal-to-noise ratio for a given bandwidth, that bandwidth being very narrow. Assuming we aren't worried about speed in the first place, it's easy to assume that very slow CW with a very narrow filter will give us something intelligible when the conditions are bad. But this theory falls apart in the case of something like the *libration fading* encountered when operating moonbounce (discussed in Chapter 16). A long dash, when subject to libration fading, can be chopped up into what looks or sounds like a lot of dots. In this case, you can actually increase your *effective* signal to noise ratio by increasing the speed to the point where a dash is shorter than (or equal to) the libration period.

The same principle can be expanded to using multitone data modes during rapid fading. You have a better chance of getting *some* information through at a high data rate, in between fades, as opposed to using a slower data rate where a symbol or character covers several fade cycles. This is somewhat akin to the Nyquist sampling theorem, where the sampling rate can be roughly approximated by the fade rate. (It doesn't *precisely* follow Nyquist, however, because the fade rates are neither uniform in time, nor completely "on" or "off.") A very crude but effective analogy would be attempting to paint a detailed picture on a very lumpy canvas. The resolution of the information you're trying to convey can't be any better than the background medium, in this case the ionospheric fading "picture."

Don't forget, also, that the ionosphere, as we've so emphatically stated, is *not* reciprocal. There may be many times where it's beneficial to operate one digital mode going one way, and another one going the other way. This would seem to be a neat area in which to do some experimentation. You don't know until you try!

Low and Slow

While the emphasis of many newer digital modes has been to increase data rates within the confines of nature, there are times when speed is the last priority. A great example would be the experimenters' band down on 500 kHz (600 meters). There's a lot of fascinating experimentation going on down there with low speed modes. By far, QRSS (Super Slow) or QRSSS (Super, Super Slow) CW are the mainstay of 600 meter operation. Some of the JT digital modes are showing promise as well. Because of the extreme power limitations imposed on the experimental 600 meter stations and the compromise nature of practical antennas for those frequencies, most signals are *well* below the noise floor, and in a region of the radio spectrum that has very high noise to start with.

QRSS/QRSSS is attractive because of its simplicity and theoretically infinite gain (assuming you have an infinite amount of time to decode the signals!). Extremely weak signals can be extracted from noise by the principle of *long time integration.* Assuming the background noise is uniform, and a zero bandwidth carrier exists in that noise, if you periodically sample everything around that carrier, and then average it, the carrier will increase with each sampling, while the noise remains constant. Signal averaging is used in many scientific disciplines, and is cheap and effective.

Another effective method, often used in conjunction with signal averaging, is *coherent detection*, which is the simple multiplication of a carrier with a known reference carrier. The simple direct conversion (DC) receiver is a coherent detector if the local oscillator is precisely of the same frequency and phase as the signal you want to detect. The bandwidth of a coherent detector is determined entirely by the bandwidth of a low-pass filter following the detector.

A special type of direct conversion receiver, called the lock-in amplifier, consists of two DC receivers operating in quadrature. The advantage of the lock-in amplifier over the simple DC receiver is that you can look at the *amplitude* of the signal, independent of the phase angle, as long as your local oscillator is within the bandwidth of the incoming signal. This may still be a challenge, however, if your bandwidth is on the order of millihertz…or microhertz. Such narrow bandwidths are easy to achieve with a lock-in amplifier. An effective lock-in amplifier requires that the receiver local oscillator and the transmitter are locked to a common reference. With the ready availability of GPS, this is not as difficult a task as it once was, especially if the transmitter and receiver are separated by hundreds or thousands of miles.

Here is a project for a homebrew GPS disciplined oscillator that can be used as a local oscillator for a coherent detector or lock-in amplifier: **www.rollanet.org/~joeh/projects/GPSlave_writeup.pdf**.

One of these would be needed at both the transmitter and receiver site. Incidentally, the use of a DDS (direct digital synthesis) module, available from several sources, allows you to use a reference frequency of any arbitrary value. So you don't need to worry about the GPS disciplined oscillator being on the local oscillator frequency itself. The accuracy and stability of the DDS will be the same as the GPS reference.

Natural Radio

No discussion of radio propagation would be complete without more than a passing mention of the fascinating world of VLF and ULF (very low frequency and ultra low frequency) radio. Although we don't have any actual Amateur Radio allocations in the VLF/ULF regions, there is plenty to listen to down there. VLF is the region between 3 and 30 kHz, while ULF is the region between 300 Hz and 3 kHz. Put together, the VLF/ULF range approximates the audio frequency range, and not too surprisingly, these frequencies can be detected using common audio amplifier technology. In fact, the typical computer sound card, in concert with a low random wire antenna, makes a pretty good VLF/ULF receiver, especially when used in conjunction with a waterfall program that allows you to see signals buried in the noise.

While there is a good deal of interesting commercial, military, and scientific communications activity in the VLF/ULF range, by far the most interesting "transmitters" are the countless natural sources that generate radio signals in that range. Signals such as "whistlers" are ionospheric VLF signals excited by distant lightning strikes. These have been observed since well before the advent of radio, and were first observed on telephone landlines, particularly military field telephone installations. Since that time, entire catalogs of new sounds such as choruses, auroral roar, auroral buzz, chirps, clickers, and more have been observed, and have raised more questions than answers.

One of the best sites on the Internet for those interested in the nether regions of radio is Radio Waves Below 22 kHz at **www.vlf.it**. This website has plenty of projects for the natural radio experimenter. One nice thing about these frequencies is that there's always something happening down there even when the HF bands are at their "deadest."

But is it Radio?

Because of the extreme wavelengths involved in VLF/ULF radio, there is a valid question as to whether what we observe is actually radio, or merely magnetic induction. In most cases, at least at

ULF, any receiver is going to be well within the *near field* of any terrestrial ULF transmitter. This ambiguity in no way detracts from the fascination of these phenomena. It simply serves to make their analysis a little trickier.

As an interesting side note, well before Marconi and his peers were “playing” with radio, telephone and telegraph hobbyists were taking advantage of magnetically coupled phone systems, using the induction to take the place of wires in many cases, over surprisingly long distances.

Closing Remarks

Modern technology has made it extremely simple to experiment with countless modes and regions of the radio spectrum once the domain of the well-heeled scientist. Don’t get into too much of an operating rut! Experiment with new modes and bands. Again, we need not remind you that if we don’t figure out what to do with our more exotic Amateur Radio bands, someone else will.

Sea Shanty

Since most of the Earth's surface is sea water, we should give maritime propagation more than a parenthetical mention. Practical radio underwent some of its most rapid and impressive development on the high seas near the turn of the 20th century. Unlike on land, where there were alternative communications technologies ranging from telegraph to telephone to carrier pigeon, before radio, ships at sea were pretty much incommunicado for long periods of time.

At the lower HF and medium wave frequencies, sea propagation is essentially ground wave propagation but with a nearly perfect ground system — the surface of the sea. Saltwater is a very good conductor, and a vertically polarized signal travels with very little attenuation over great distances along its surface. Experienced ship telegraph operators will tell you that the 500 kHz frequency maritime radio "network" was a 24 hour a day party line, with extremely reliable communications to and from any ship on the high seas to any other ship on the high seas.

Of course, 500 kHz is not available for radio amateurs, except on a limited experimental basis, but 160 meters (1.8 MHz) can offer a lot of the same benefits for the seafaring radio amateur.

Making Things Plane

As we have described before, normal radio waves in free space attenuate because the rays expand in a spherical fashion from the source. This results in a field voltage attenuating as 1/R.

Because a sea-surface wave is guided, the pure *geometric* attenuation should be much less. Rather than the waves spreading out in a sphere, they spread out more or less in a planar fashion, attenuating only in two dimensions rather than three. The real-world experiences of our aforementioned ships' telegraphers seem to bear out this assertion. While there is some resistive loss of a surface conducted wave, the geometric attenuation is so much less than in free space that we should expect some extremely reliable operation over the sea's surface.

While the majority of the surface of the Earth is seawater, a relatively small number of radio amateurs actually operate on the sea, or even communicate with ships at sea. The real significance of sea propagation is that a very large percentage of DX Amateur Radio communications take place over maritime paths. Most long Amateur Radio propagation paths make at least *some* of their journey over one or more seas, which makes understanding sea propagation all that more interesting.

What's for Launch?

Nearly every radio amateur who has had the privilege of operating near the seashore will tell you that DXing just works that much better when you're near the sea, and that when a signal has a "choice," the propagation works best in the direction of the most water. This observation is not limited to those low bands where surface propagation can be taken advantage of; this observation seems to hold even on the high bands where ground wave propagation is all but impossible.

What is happening here? Is there something that favors radio propagation over the sea, even when a surface wave is impossible? Does being near a big body of sea water help launch the signal in some way?

As it turns out, there are a number of factors at play here. It's probably more helpful to talk about the immediate environment of the antenna and work outward to get the complete picture.

Groundwater

As we've described in the first few chapters, most hams severely underestimate the contribution of the ground in the vicinity of their antenna systems…either for good or ill. Unless you have a really tall tower on 10 meters (or higher frequency bands) the ground under your antenna is going to be pretty tightly coupled to your antenna. In addition, the character of the ground a significant number of wavelengths away from the base of your antenna is going to have a significant effect. Only in recent years have antenna modeling programs really taken the "far ground" seriously.

For the sake of argument, we can consider a 20 meter horizontally polarized Yagi about 1⁄4 wavelength above ground, right on the beach of a large continent (as opposed to being on a rock in the middle of the ocean). The continental "dirt" is of only moderate conductivity, forming an average ground. In the direction of the water, we have very good conductivity. The overall arrangement would be similar to having half a buried radial system, with the existing radials "aiming" out toward the ocean. We know that the overall beam shape is a function of both the antenna's free-space pattern, and that of the ground reflection. A perfect ground underneath a horizontal antenna (either a simple dipole, or a gain antenna) will give you the lowest possible launch angle, all other things being equal. This can be confirmed by looking at the catalog of patterns versus height above ground in any antenna reference such as the *ARRL Antenna Book*. So we know we're going to have a nice low launch angle out over the ocean, which is generally desirable for good DX. Looking in the landward direction, it's likely that our effective height above ground is going to be a bit higher, as the reflection surface of average "dirt" is well below the surface. Plus it's generally a lot lossier. So we have a less reflective ground at a less desirable height over the continental path.

Well, that all makes perfect sense with regard to launching the signal in the first place. But what about after a few skips where the initial radiation pattern is irrelevant? Is there reason to believe that a radio signal propagates better over the ocean than over land, again disregarding vertically polarized surface waves?

The answer would be a qualified yes. Not that propagation over sea water, *per se*, is less lossy; but rather that reflections from the sea's surface are going to be less lossy than reflections from normal earth. The more skips you have, the greater this difference would be noticed. The reflecting surface of the sea is well defined, and as a result, reflections therefrom are going to be nearly specular (mirror-like). Reflections from normal ground are not only going to be lossier, but the depth of penetration over a number of hops is more than likely going to be very inconsistent, resulting in a rather diffuse (and lumpy) reflected wavefront.

Naval Navigating

In general, radio amateurs aren't primarily interested in radio navigation, but we can learn a few things about radio propagation by looking at some classic issues of the technology.

Radio direction finding (RDF) is one of the disciplines first developed on the high seas. Again, like radio itself, RDF began on the low end of the radio spectrum. Vertically polarized medium wave signals travel with low attenuation in contact with the surface of the sea.

The traditional means of homing in on such signals is by means of a vertically polarized *tuned loop*

antenna. The tuned loop consists of a number of turns of wire in a loop a foot or two in diameter, preferably enclosed in an electrostatic shield. The ferrite loopstick antenna built into most AM portable radios has similar characteristics to the classical tuned loop, minus the electrostatic shielding. If the tuned loop is very small relative to the wavelength, it exhibits the properties of a *wave antenna*…
very different from a full sized quad loop or delta loop element. See **Figure 19.1**.

The tuned loop antenna is very strongly vertically polarized, and has a very sharp null perpendicular to the plane of the loop. (Actually two nulls, 180 degrees apart). For true sea-surface waves, this antenna is extremely precise, and was used for many decades to home in on shipboard or land based beacon transmitters by swinging the antenna around for a sharp null.

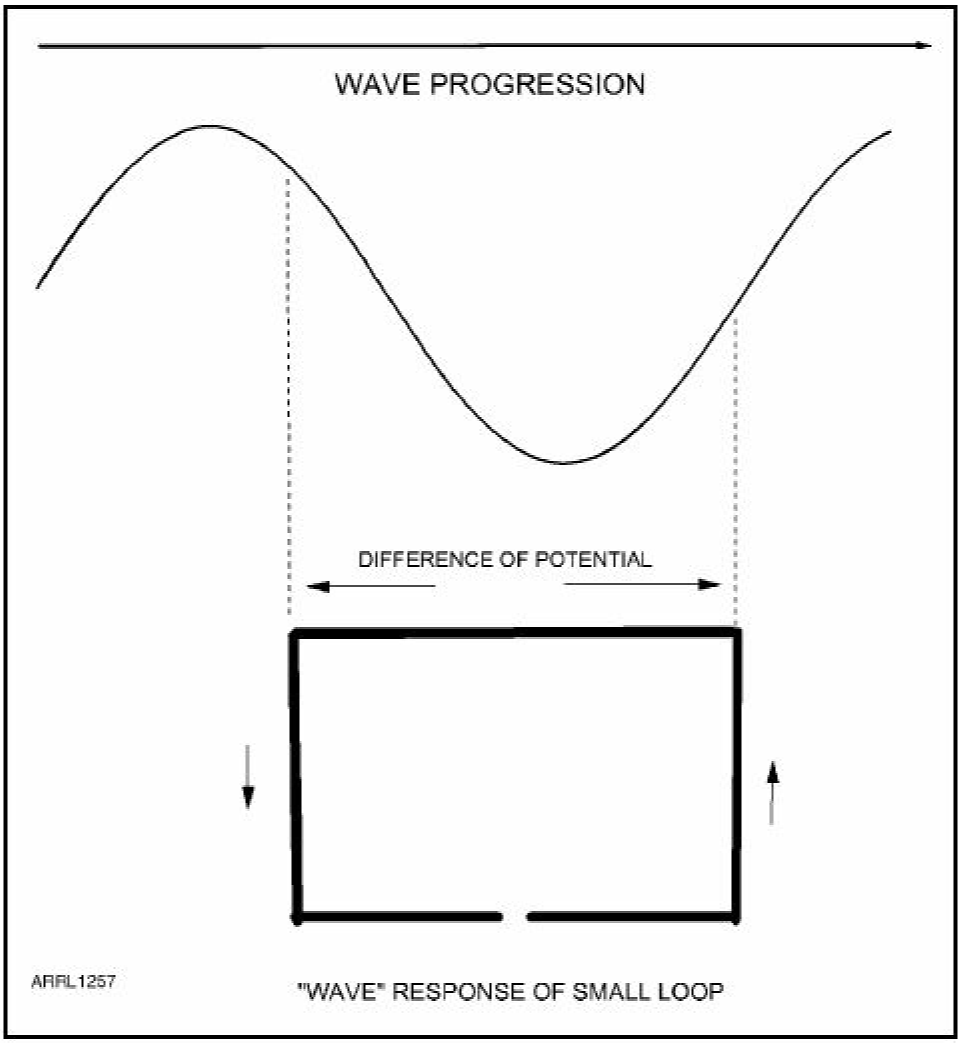


Figure 19.1 — A wave antenna operates in a rather different fashion than a more typical receiving antenna. As a wave progresses through two different regions of the antenna, different voltages are induced because of the difference in phase of the passing wave. The sensitivity of such an antenna is directly related to the physical size of the antenna, as a greater phase difference will be seen in the larger antenna. This principle applies whether one is talking about a small tuned loop or a long Beverage antenna.

However, the unshielded loop can suffer from serious error if signals arrived from high angles. While a vertically polarized antenna can launch a good "sea wave" there is nothing to stop it from also launching a vertically polarized skywave. During the daytime, D layer absorption all but eliminates skywave propagation. During the night, when the D layer absorption is low or nonexistent, skywave signals can arrive in addition to the desired "sea wave."

The first undesirable effect of this is that the user will not be able to obtain perfect or even reasonably sharp nulls with the loop. For signals arriving at angles above the horizon, the wave antenna properties of the loop no longer fully apply, and the entire loop can act as a short whip antenna. This is called “night effect,” naturally enough, because it only occurs at night. This is the primary purpose of electrostatic shielding, as it prevents the antenna from working in any but the desired *wave mode.*

Contrary to a lot of erroneous literature, the electrostatic shielding on a tuned loop does nothing to reduce static pickup. The truth is that electrostatic shielding helps the antenna maintain its proper pattern, which *can* assist in nulling out local noise. But again, this is not because of any “noise shielding” that supposedly exists in such an antenna.

In addition to the skywave possibly spoiling the nulling capacity of a loop antenna, it can, to a large degree, also displace the null of even the best shielded and balanced loop. As we discovered, skywave signals can be skewed both in elevation and azimuth by the Earth’s magnetic field. This effect can be quite dramatic at medium wave frequencies. While the sea wave signal is compelled to follow a true path along the sea surface, no skywave has the same constraints. So, unless you know you’re really looking at a sea wave, you can be in for some serious errors.

The Adcock antenna was the next level of development for reducing skywave interference to the direction finding process. The Adcock antenna is a modified form of phased array which is capable of a null at very low angles of radiation. Unlike the small loop *wave antenna,* the Adcock is more traditional in operation, and is capable of effective operation on higher frequency HF bands, as well.

This brings up another very interesting point about sea propagation. As we described in our initial discussion of ground wave propagation, generally HF signals are attenuated rapidly, and thus only low band or medium wave signals are suitable for true ground wave propagation. This restriction is nowhere near as dire in the case of sea wave, however. HF signals up to 15 MHz or so can indeed travel with relatively low loss along the surface of the sea! The attenuation is certainly greater than that of free space, but can still result in some very long distance sea-wave propagation.

This is one factor that probably accounts for the many glowing reports of DX by hams using shore-based vertical antennas. Certainly a vertical antenna has inherently low launch angles, but this probably is not the entire story. The one way to absolutely confirm this would be to establish HF continual communications with a ship traveling away from shore for a few thousand miles and see if there are any dead spots (skip zones). This could be an interesting project for experimentally minded hams.

NVIS Modes and Methods

In recent years, radio amateurs have taken a renewed interest in emergency communications, especially those which don't require a lot of high technology or fragile infrastructure. Closely connected with this development is the renewal of interest in *near vertical incident skywave (NVIS)* propagation and methods. While NVIS is as old as Amateur Radio, the acronym itself would be difficult to find in any amateur literature more than a decade old or so. More traditionally, this was referred to as "short skip" or "really short skip."

Most practical Amateur Radio emergency and public service communications take place over moderate distances, generally no more than 100 miles. The same applies to regional traffic nets and even ragchew nets. While low band ground wave is effective and reliable for perhaps 50 miles or so, at 100 miles or so it's at the outer limit of effectiveness for most true ground wave. (Take note of "sea-wave" exceptions in the previous chapter).

Up and Down

A perfect *vertical incident skywave (VIS)* signal would only be receivable at precisely the location from which it was launched. Here's an interesting little point you may not have thought of before. If the Earth had a diameter of zero, and the ionosphere were a sphere of, say 5000 miles in diameter, *all* signals would be VIS signals if launched from the Earth, no matter which way you aimed them! The reality is the outer surface of the Earth is nearly the same diameter as the outer reaches of the ionosphere. This is where a lot of the models of the ionosphere you've encountered are a bit skewed. If you were to expand the surface of the Earth to the size of an orange, the surface of the Earth would be the inner surface of the rind, and the upper reaches of the F2 layer would be the outside surface of the rind! So, relatively speaking, we can consider the "ceiling" to be a horizontal plane, rather than a sphere.

Now, the vertical incident ionosphere we discussed theoretically looks at only true VIS signals. In practice, any physically "doable" antenna has a significant beamwidth, so even if it's aimed precisely vertically, it has a combination of a VIS ray (dead center of the pattern) surrounded by a cone of NVIS rays. Only the true VIS ray in the middle of the cone is going to come precisely straight down, while most of the cone's rays come down at some distance from the launch site.

NVIS signals are generally very low loss, because they spend relatively little time in the ionosphere; in other words, their total paths are relatively short. Effective NVIS can occur at any frequency below the critical frequency but above the D-layer absorption frequency (LUF).

In practice, the best frequency is about 10% below the critical frequency. Obviously there are a lot of times when this particular optimal frequency won't fall within any ham band, so it's best to choose the highest frequency ham band that's below the critical frequency in your region.

On rare occasions, the critical frequency can actually be below the LUF, in which case, NVIS propagation would be impossible. In reality, NVIS on the low bands is useful nearly from sundown to sunrise,in most locations. It takes very little ionization to create effective reflections on 80 meters.

Recall that as the local evening approaches, the ionosphere retreats from the bottom end up, meaning the band will “go long.” This is particularly noticeable on NVIS links, where even very low antennas may become very effective for DX!

Cloud Burners

Hams call any antenna that concentrates most of its energy straight upward a “cloud burner.” This does not mean that NVIS requires high gain antennas. However, if you really want to optimize your station for excellent NVIS reliability, there are some effective favorite antennas. One of the best low-band cloud burners is essentially a cubical quad lying on its back. It is a full wave loop, about 1⁄4 wavelength above ground, with a full wave loop reflector at or slightly above ground level. This is a very popular antenna in Alaska, where NVIS dominates the regional evening nets on 75 meters. These antennas produce consistently imposing signals just about anywhere in interior Alaska, usually several S units above those stations using verticals, and even inverted Vs and dipoles.

In the absence of the motivation (or real estate) for such an antenna, a very effective alternative is a dipole or inverted V at 1⁄4 wave with a reflector mounted underneath. While a large reflective ground screen is ideal, one or two reflecting wires can gain you a dB or two of gain over a dipole. Another effective arrangement is a phased array of two or more dipoles operating in a broadside manner. This will also allow you to steer the antenna to some degree along one axis, by adjusting the phasing, which may help at times with NVIS communications.

X and O Once More

Probably nowhere will you notice a more profound difference between X and O propagation than when using NVIS. One very effective NVIS antenna is a pair of crossed inverted Vs, fed 90 electrical degrees apart. This is the standard configuration for ionospheric research, where such antennas are often combined in large 2D arrays. This antenna only needs one tall support, such as a push up pole, with the antenna elements serving as the guy wires. By switching the phasing of this antenna you can select either X or O mode propagation. You can easily observe a three S unit difference in signal strength between X and O modes.

In addition, by selectively *transmitting* through a CPOL antenna, you can select between two reflection heights in the ionosphere, thus changing the range of your signal. This is a method that is just beginning to gain some traction in interior Alaska, and presumably some other areas that rely heavily on NVIS.

X and O mode selectivity can be built into any type of antenna, though it can take up a bit more real estate for some designs. Two of the aforementioned quad loop cloud burners fed in quadrature would make a killer X-O NVIS antenna. One of the interesting aspects of circular polarization is that the N-S and E-W axes don’t need to cross each other. Two quad loops side by side, fed in quadrature, make a perfectly acceptable CPOL antenna.

Billboard

One of the most surprising and interesting phenomenon one might encounter when using an NVIS antenna, especially in high latitudes, is how well it can work as a DX antenna…at times. If a signal is launched nearly vertically and encounters a horizontal reflector (the ionosphere) the signal needs to return nearly vertically as well. The big *if* here is if the ionosphere is horizontal! As it turns out, the "horizontalness" of the ionosphere is a broad assumption that falls apart in many places at various times. In polar or near-polar regions, the ionosphere can slope down severely, causing it to appear as a big slanted billboard or drive-in movie screen (**Figure 20.1**). NVIS signals can reflect off this tilted layer and be reflected at horizontal or near horizontal angles, forming a very effective DX antenna.

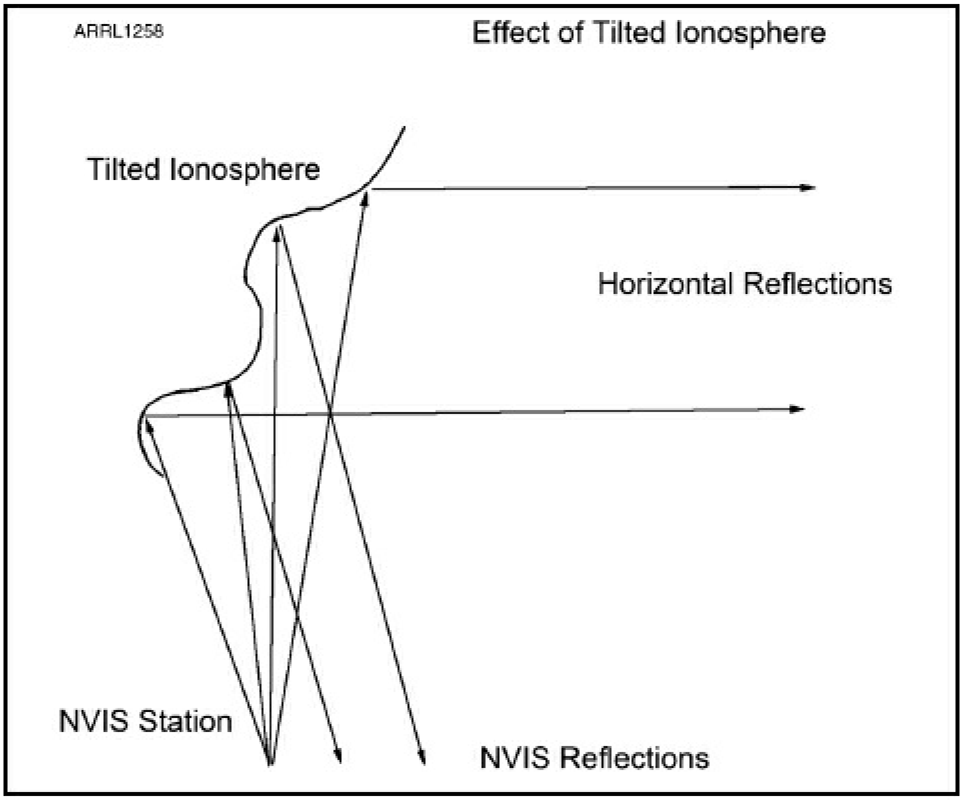


Figure 20.1 — A tilted ionosphere clearly explains why an NVIS signal at one location can appear as a low angle signal at another, and vice-versa. The majority of propagation prediction software packages do not account for ionospheric tilt. *Proplab Pro*, described in Chapter 15, is one notable exception.

In Chapter 15, we mentioned the propagation prediction software package *Proplab Pro*. In addition to being the only readily available software package that deals with X and O propagation, it is also the only program that deals with ionospheric tilt. Both real-life experience and this modeling program confirm that this ionospheric tilt can be very dramatic, up to 60 degrees or more near the Arctic Circle! This can certainly put some kinks in our understanding of "normal" ionospheric propagation.

Like all ionospheric phenomena, ionospheric tilt is subject to change. During much of the year the ionosphere over interior Alaska can be pretty normal, which is why NVIS acts like NVIS a good part of the year. But when ionospheric tilt is in play, propagation can be phenomenal…as well has highly non-reciprocal.

Interior Alaska is often referred to as the "Bermuda Triangle" of radio. While signals from just about anywhere can be heard most of the time, there are large chunks of the year when nothing at all seems to get out. It's not that the absorption is excessively high in the region, but rather that the Earth's magnetic field is nearly vertical at this location, which has profound effects on the divergence of the X and O waves, neither of which go in expected directions. This factor combined with ionospheric tilt can make radio propagation so difficult to predict as to be deemed non-existent.

The advent of *Proplab Pro* begins to make some sense of our strange northern propagation. Incidentally, most of what is experienced in high northern latitudes should be similarly observed in high southern latitudes. However, since the population is so much lower in the southern polar regions, these phenomena aren't as well documented or complained about!

60 Meter Magic

No discussion of NVIS propagation would be complete without some discussion of 60 meters. 60 meters is actually a very good band for NVIS, since it is nearly always above the D-layer absorption frequency, and yet almost always below the critical frequency, regardless of your location. This discovery was made in interior Alaska long before we had an amateur band in that frequency range. The Alaska public fixed frequency of 5167.5 kHz (actually a maritime allocation) has for decades been a default emergency frequency in Bush Alaska. Although regular Amateur Radio use of this frequency was forbidden, regular equipment tests were permitted, and this frequency proved to be very reliable for most of the state. So when 60 meters first became available, many Alaskan hams knew just what to do with it. It is highly recommended that emergency communications minded hams fully explore 60 meters for its reliable regional NVIS characteristics.

Also, as in the case of some of our other bands, having regular activity on the band is more likely to elicit some DX responses as well. On occasions too numerous to mention, many ham bands are deemed to be dead, when the fact is simply that nobody is on the air!

NVIS Mobile Tricks and Treats

It's probably fair to say that the military has led the charge for mobile NVIS communications. An Army communications manual from the 1950s illustrates a 12 foot "tank whip" mounted on the rear of a jeep, bent forward with its top end secured to the front bumper by a short rope to *enhance high angle propagation*, for more reliable local communications. There was little else said about the mechanics of skywave propagation except to recommend that this method works. (As with most military equipment manuals of the era, they seldom offered the reader much more than he or she needed to know).

Well, the fact of the matter is that this does work, and it works quite well. What the Army lacks in verbosity, they make up for in pragmatism.

Strolling through the parking lot of the Dayton Hamvention, one will encounter a veritable forest of *serious* HF mobile antennas. Many of these are low band "Texas Bugcatchers" and such, optimized for effective ground wave coverage, no mean feat considering the restraints one has to work with. Building an HF mobile antenna that works at all takes some serious plotting and planning, not to mention large loading coils with a lot of wind drag.

As we've already well established, low frequency ground waves attenuate rapidly, even if you have a good ground to start with — a privilege you don't have in any mobile installation. Optimizing a mobile antenna for NVIS can significantly reduce the hardware requirements for effective HF mobile operation.

First of all, an NVIS antenna doesn't need to be anywhere near as *efficient* as a typical HF mobile antenna. The path loss of an NVIS signal is much less than ground wave for most low-band HF frequencies and situations. In most cases, with a long whip, installed as in the Jeep example above, a loading coil or tuner at the *base* of the antenna is more than adequate. Certainly a center loaded whip, such as a Texas Bugcatcher is much better from an efficiency standpoint. However, using NVIS, we aren't trying to launch a worm-roasting ground wave.

Now, the astute Amateur Radio antenna guru is certain to remind us that bending over an antenna so that a large portion of it is parallel and in close proximity to the car's body will lower the radiation resistance and efficiency. This is true. But the gains achieved by taking advantage of NVIS still outweigh this increased loss in many if not most situations.

Here's another trick you may not have thought about. Instead of just using a rope to hold the "top" of the antenna down near the front bumper, you could insert a capacitor between the tip of the antenna and the car body. This capacitive end loading will create a much more uniform current distribution on the whip, raising the effective radiation resistance in the process. You will find that you need less loading inductance at the base, as well. Capacitive loading is almost always more efficient than inductive loading, anyway.

Stationary Mobile NVIS

A mobile "station" may be a contradiction in terms, because by definition a station is… well…*stationary.* Nevertheless, there are many instances where a mobile installation is very effective even when you aren't moving. In light of the fact that operating a radio while driving is being looked upon in ever more suspicious terms (perhaps rightfully so), it will be worth our while to look at "parked mobile" operations, especially of the NVIS variety.

While folding a rear bumper mounted whip forward decreases the radiation resistance, folding it *backward* does the opposite; it increases the radiation resistance and efficiency. Needless to say, driving a vehicle with an 8 foot whip poking out straight backward is probably frowned upon, but this can be an effective NVIS antenna while parked at your friendly local rest stop or even in your driveway.

The Radiating Car

Despite the fact that HF mobile antennas have been discussed in great depth for decades, there's one aspect that has been somewhat misrepresented. Most models and discussions of HF mobile operation describe the vehicle's body in terms of being an amorphous capacitor plate, with the Earth beneath being some kind of counterpoise. This *might* be a semi-valid analysis if the vertical antenna were precisely in the center of the body of the car, at frequencies where the car body is on the order of a half wavelength. This would approximate a "two radial" ground plane of sorts. Of course, this model would be entirely invalid for low bands where the vehicle is far too short in any direction to be a "radial" as well as having far too little surface area to form a counterpoise.

Recent, and not so recent, antenna modeling has long disproven this counterpoise concept. We strongly recommend you read the late L.B. Cebik's excellent treatise on the counterpoise — and how no such device actually exists! (It's published online at **www.antennex.com/shack/Dec06/cps.html**.) The truth of the matter is that the typical rear bumper mounted low-band HF antenna appears as a vertical whip with one very short radial. The fact that the car's body has a lot of mass or volume has zero relevance with regard to radiation. The entire car body can be resolved into a single wire, moving away from the base of the whip, toward the front of the vehicle. This can be proven with any valid *NEC* modeling program, and confirmed with near field current measurements.

Now, while this "radial" is very short in terms of wavelength (at least on the low bands) it is roughly of the same order as a typical vertical whip — eight feet or so. What this means is that in actuality about half the signal is radiated from the whip and the other half from the car body. This can be viewed as an inverted V lying on one leg. The main lobe of radiation bisects the V. In the mobile installation, this means the radiation is at about 45 degrees above the hood of the car. (Again, this can be confirmed by actual field strength measurements).

Having established that the car body in most cases radiates just as much as the mobile antenna itself, if we were to fold the antenna back horizontally, we can reasonably approximate a center loaded, very short dipole (again, assuming low band operation). And, of course, this dipole would be ideally oriented for NVIS operation, though perhaps, not at an optimal height.

For Further Investigation

While not directly related to NVIS operation, if we recognize that a typical vehicle is not a counterpoise, but rather half of an antenna, we might consider doing things a bit differently. For example, in the case of base loading, it might be beneficial to split the loading coil in half, one half serving as the loading coil for the car body, and the other half being for the antenna. The RF would be fed to the center of this coil, instead of at the bottom end. This might be a little more complicated to implement from a circuit topology standpoint, but it may conceivably increase the overall efficiency by “loading” the car itself.

A Viable Alternative

As more and more hams face restrictive antenna situations at their home stations, the effective HF mobile installation is a very attractive alternative. In fact, in many cases a mobile ham station can be much better than a home station even without considering antenna restrictions. Even in the most urbanized areas of the country, you can often drive to where manmade noise levels are nearly nonexistent. Being NVIS savvy can only add to your mobile Amateur Radio enjoyment and effectiveness.

Unexplored Territory

Amateur Radio has been around for a century now, and while we have learned a great deal about how radio works, there is still a lot we don't know. And there's no indication that we will ever come to the end of the road when it comes to radio know-how. The reason a good number of us are hams is to satisfy our curiosity as much as to provide practical communications. Of course, these are not mutually exclusive pursuits by any stretch of the imagination. Most of the technologies and techniques we consider absolutely necessary were at one time the result of some experiment having gone seriously awry. Any student of the history of science knows this to be the case. Scientific knowledge is seldom an orderly, linear path.

We trust this volume has imparted enough basic knowledge to spare the newcomer the need to reinvent the wheel, as well as to help explain a few of the behaviors even the more experienced ham may never have considered. Knowing how radio is *supposed to work* will save hours of frustration for the new radio amateur. The science and physics of radio propagation are very different from the technology used to exploit them. Most new hams "get" technology, while relatively few of them "get" radio propagation.

At one time, many radio amateurs arrived in the hobby with a fairly good feel for radio propagation, especially on the HF bands, having had some exposure to shortwave radio. It is safe to say that the vast majority of new Technician licensees (and many other amateurs as well) have had no exposure whatsoever to shortwave radio, and have no idea what to expect before they get on the air… especially on the HF bands.

It's a good thing to know as much as possible of the fascinating subject of radio propagation before getting on the air. But perhaps too much knowledge might not be a good thing.

Avoiding the Paralysis of Analysis

Knowledge is supposed to make things easier. As we suggested at the beginning of this tome, it is truly amazing what early radio amateurs were able to accomplish with very little understanding. While we want to give the current generation of hams a running start over those early pioneers, we also want to avoid going too far in the other extreme. Countless newcomers to Amateur Radio are convinced that if they don't have a high-end transceiver and a sky-filling multiband Yagi atop an Eiffel Tower they can't do effective long distance radio communications. A disconcerting number of hams never progress beyond their Technician licenses (or even use all of the privileges they have *within* that license) because they have the erroneous impression that HF radio is a big ticket item. We old timers need to remind them (and ourselves) that for many decades, radio amateurs worked the world with rudimentary homemade transmitters feeding what our friends across the Pond refer to as "tatty bits of wire."

This can be summed up by saying that Amateur Radio is for experimenters. This book is intended as a guide, not the final word on the topic. Practical and theoretical knowledge gleaned from this work should instigate further exploration and discovery; it is this aspect that will keep Amateur Radio vibrant and fascinating for the foreseeable future. While we have touched on most of the known topics pertinent to radio propagation in all its forms, there are still aspects where we are in our collective infancy with regard to our understanding. We will close with an overview of subjects that still need explanation and exploration.

Antennas

While antennas and radio propagation are entirely different topics, they are inextricably linked. Our only "window" into radio propagation is the antenna. It is the only interface between electromagnetic radiation and our physical senses. While the interaction between antennas and electromagnetic fields is well understood, there's still a lot of work to be done with antenna designs that truly take advantage of all the peculiarities of radio propagation.

In particular, amateurs can do more with circular polarization at HF. Although CPOL has been explored in the past, it has not been widely accepted. Carl Leutzelschwab, K9LA, points out that in the December 1962 *RSGB Bulletin*, George Messenger, the original holder of the K6CT call sign, had an article noting the superiority of crossed Yagis fed in quadrature. He formed the Space-Raider antenna company, offering CPOL 6, 10, 15 and 20 meter antennas with ads in *QST* magazines of the era. Unfortunately they never caught on. In the November 1989 issue of *Practical Wireless* (later reprinted in the November 1990 issue of *Communications Quarterly*), B. Sykes, G2HCG, described a CPOL 10 meter antenna (again using two Yagis fed in quadrature) and noted observations on 10 meter beacons.

Band Usage

We have a lot of Amateur Radio bands at our disposal, each of them with very different properties. We have 12 microwave bands that beg for more exploration. Each of these microwave bands has unique properties that aren't likely to be discovered until hams do more with them! Microwave transmitter and receiver design is not for the faint of heart, but the rewards can be huge. Exploration of the microwave bands is one area where hams can still make a major contribution to the radio art. We dare not take these valuable allocations lightly!

Coding

We've discussed the plethora of novel digital modes available to the radio amateur. However, there are undoubtedly many new modes to be discovered. Inexpensive and powerful tools such as Arduino, Raspberry Pi, and other similar tools put previously daunting computer coding into the hands of just about anyone. In a recent discussion with an Amateur Radio equipment manufacturer, we learned that they are interested in creating an ionosonde that works with FCC Part 15 power levels… all because of clever signal processor coding. A device like a Part 15 ionosonde would revolutionize Amateur Radio ionospheric research.

Digital Relay Stations

There's still a lot to be learned with regard to effective digital relay technology, such as Winlink, ALE, and a number of other automatic relay modes. A lot of this has more to do with regulatory issues than technology, but it's nevertheless a wide open field for the experimenter.

Experimentation

Of course, this is all-inclusive. There are certainly propagation modes we haven't even discovered yet. We are likely to discover these on the microwave bands, but we could also discover new things down on the VLF bands as well.

Field Aligned Irregularities

Field aligned irregularities (FAI) events are structures in the auroral zone. These are thought to be formed by "plasma tubes" that follow the magnetic field lines through the aurora. These structures are highly reflective, though quite "bumpy." Effective VHF and UHF communications can be achieved by using FAIs. Experimentation with more effectively using FAIs would mark a significant new technology

Ground Wave Propagation (Sea Wave Propagation)

Although ground wave propagation is as old as radio, there are still aspects of it that remain a mystery. "When all else fails, use ground wave" may be a valid philosophy if we have effective means of taking advantage of ground wave signals. Modern digital signal processing technologies should enable ground wave communications under conditions once deemed unusable.

Heating

Ionospheric heating is normally deemed to be the domain of super high powered transmitter facilities such as HAARP and EISCAT. Ionospheric heating is the method of "pumping" the ionosphere into nonlinearity, where all kinds of interesting phenomena can occur. Recent observations, however, indicate that some ionospheric heating can occur at Amateur Radio power levels. This is wide open for experimentation.

Ionospheric Physics

Another all-encompassing topic. There's a lot we know about ionospheric physics. And there's a lot we don't know. Fully understanding ionospheric physics goes far beyond radio; it could even lead to practical controlled nuclear fusion.

Weak Signal Digital Modes

The JT weak signal digital modes have opened up entire new worlds for the radio amateur. We certainly have not exhausted the possibilities of such weak signal modes. We should probably expect to see many similar JT modes in the near future.

K-index and its Kin

There's still a somewhat tenuous connection between solar indices and radio propagation. For example, 160 meters seems to defy a lot of conventional wisdom when it comes to solar activity and radio "goodness." While we certainly cannot ignore solar indices, they far from explain all that we observe when it comes to ionospheric radio propagation. More powerful computing methods may make the connection more clear.

LEOs other Launches

Low Earth orbiting satellites can reveal a lot about the ionosphere and atmosphere. Radio amateurs have had a great deal of input when it comes to scientific satellites, and we need to continue to pursue this. Even high altitude balloons can reveal a lot about our radio world.

Magnetic Fields

As we've discussed in detail, the Earth's magnetic field has been all but ignored in radio propagation studies, at least in Amateur Radio circles. Now, there is no excuse to neglect this all-important ingredient of radio propagation, especially at HF In addition to propagation prediction programming, considering the Earth's magnetic field will change the way we do radio.

NVIS

We suspect that certain digital modes may be particularly suitable to NVIS operation. In addition, the development of compact NVIS antennas can solve a lot of antenna restriction issues. Antenna experimentation and construction will always be part and parcel of Amateur Radio. NVIS is one area where there aren't *yet* a lot of original ideas. This is a wide open field.

Open Source

While not directly related to propagation, every ham needs to know that software is the "new homebrew." There is so much wonderful open source software and hardware out there, that there is little excuse for any ham not to "roll your own." The Open Source movement is closely linked with the Amateur Radio philosophy.

Polarization Modulation

One of the more curious modes available to the radio amateur (though not explicitly spelled out in the FCC rules) is polarization modulation. Polarization modulation can include incremental shifts in wave polarization in response to voice modulation, or it can involve shifting between horizontal and vertical polarization in response to a MARK or a SPACE. In theory, polarization modulation could transmit intelligence with zero bandwidth. In practice, the transmission of extremely high bandwidths with very little bandwidth should be possible.

Quadrature Methods

As we described in the first chapters, the entire universe seems to be based on orthogonal axes in time, space, and even frequency. While quadrature modulation is common in many commercial and amateur applications, it could also be applied directly to electromagnetic radiation, such as digital polarization modulation described above. Or it can involve switching between right and left hand circular polarization at a digital rate. While in theory, these methods should have very predictable results in free space, they may have quite unexpected results in the ionosphere or other dispersive media. You don't know until you try it!

Radio Sounding

Included in this we could include beacon transmitters, channel probes, and ionosondes. While prediction of propagation is important, it's just as important to know what propagation actually *is*, rather than just what is should be. Radio amateurs miss a lot of band openings — from dc to daylight — because nobody happens to be listening. We need to start looking at putting more beacons out there, especially on the microwave bands. This could also encourage a lot more activity.

Signal Measurement

We covered this topic somewhat extensively in the chapter on S meters, but there's still work to be done. We need some good inexpensive field strength meters, as well as more amateurs who understand the importance of accurate signal measurement. This one discipline could go a long way toward making Amateur Radio a more scientific hobby.

Transequatorial Propagation

No matter how you cut it, transequatorial propagation is still *weird*. While a number of "explanations" for TE have been put forth over the years, including the ones already described in this book, none of them are fully satisfactory. The true mystery of TE will probably yield only the careful application of 3D triangulation described earlier. This method will tell us precisely where the reflection points are, rather than mere speculation.

Unidirectional Propagation

While the mechanisms of one-way ionospheric propagation are well understood, one way propagation has been observed on every band from dc to daylight. While some of the processes are probably analogous to ionospheric phenomena, we suspect there might be a few new things to learn.

Very Weak Signal Work

JT-65 and other digital modes have brought us a long way in this regard, but there are still a few tricks to be learned. High tech isn't necessarily synonymous with "digital." There are still some impressive analog methods for making sense out of chaos. The lock-in amplifier is one such method, as described in an earlier chapter. New low noise devices, some silicon, some not, issue forth from the chip makers on a regular basis.

Wave Antennas

This special class of antennas deserves a lot more research. Wave antennas include devices as varied as the venerable Beverage to the multiturn tuned loop. Wave antennas reveal some basic and not so basic properties of electromagnetic waves, especially in inhomogeneous media. For example, a wave antenna can be used to determine the velocity factor (and dielectric constant) of the Earth or other media.

X-Mode Propagation

While O-mode propagation is generally the most observed mode, primarily because it propagates with less loss, the X-mode ray can reveal a lot more about the ionosphere and the related physics. It is certainly gratifying to see the amateur literature finally recognize the existence of X mode propagation (as well as O-mode propagation), there's still a long ways to go in this regard. We need to keep driving home the point that HF propagation *is* either X or O mode.

Youth

As fascinating as this whole radio propagation business is, most of the world's experts, both commercial and amateur are *old people.* And yet the need for radio propagation experts has never been greater. Where are the experts designing all the next wireless devices going to come from? Radio science is actually very fun stuff, and we need to convey that fact to as many young scientists as we can. Radio technology is as central to our dominance in the science and tech fields as digital design. In fact, at the higher speeds, even digital technology becomes an RF problem!

Zulu Time

Well, perhaps not so much Zulu time, but time in general. The emergence of the GPS receiver has made radio science a whole lot easier and more accurate. We already described the GPS receiver used as a frequency reference for a lock-in amplifier. Better time and better instrumentation all around makes the study and understanding of radio propagation all that much better.

Your Own Path

Undoubtedly we have left out a lot of possibilities. As a radio amateur, an insatiable curiosity will be your best tool for advancing the state of the radio art. Radio allows the "Average Joe" access to the mysteries of the universe that no other activity can remotely match. Everything we know about the universe outside of our own back yards is revealed to us through electromagnetic radiation…radio propagation. We trust this book has opened a small window to that vast universe.

The Romance of Radio, or The Art of Being There

You can have a successful Amateur Radio experience knowing little or nothing of what has been explored in this tome. For more than a century, countless hams have made untold millions of exciting and rewarding radio contacts using nothing but "tatty bits of wire draped over the willows."

There has always been the danger (probably more imagined than real) of sacrificing the romance and intrigue of radio, particularly of the amateur sort, by understanding it too well. For generations of radio practitioners, both amateur and commercial, the unpredictability and fickleness of radio has been its main attraction. If you are a typical radio amateur, there is always the sense that there is something more to learn, some new quirk to discover.

Of course, there are a number of inviolable laws of nature we must work (or play) under. Maxwell's equations are chiseled in granite, or perhaps more aptly, the fabric of space itself. We know that even the most bizarre, fleeting, and sometimes disturbing radio phenomenon we might encounter will eventually be found to yield unswervingly to some law. This is not to say we *know* all the laws. And even if we did know all the applicable radio laws, we know that not all of them apply all the time in equal doses. Radio propagation, except in empty space, is always a complicated process, as we've learned.

Having a deep knowledge of all the applicable laws doesn't necessarily give us any more dominion over the subject, either. Knowing exactly how sunspots affect ionospheric behavior doesn't give us one iota of advantage over the neophyte, if we can't *do* anything about sunspots.

Furthermore, all the predictive tools we now have, which our recent radio ancestors could only dream of, can be a mixed blessing. The more we understand about radio, the more unlikely it seems that the thing can work at all, .at least in a *real* environment. When it comes to propagation modeling, a detailed analysis will, as often as not, tell us it's not even worth bothering to turn on the rig. This is, indeed, the *real* danger of "too much" knowledge in this area. It applies in a few other spheres of life, as well. The "paralysis of analysis" is a very real phenomenon, and it can be insidious.

We cannot over-emphasize the importance of just being on the air…a *lot*. Radio openings that defy the rules can occur at any time, any place, and from basically "dc to Daylight."

Of course, we need to acknowledge that in many cases, the time one can allocate to Amateur Radio is limited. Amateur Radio is, after all, *amateur*. A fortunate few among us are able to parlay their Amateur Radio passion into a career, where they can spend a good part of their waking hours playing with radio in some incarnation or other. But this probably does not describe the typical reader.

The judicious use of propagation prediction tools *can* make our limited, precious hours on the air more fruitful, in the same way that a good antenna modeling program can save you a lot of cutting and pruning and trial and error out in the field or antenna range. But prediction modeling is not a substitute for experience. No computer program can give you actual *insight,* or suggest a new avenue for exploration. The latter is, more often than not, the result of sheer luck…being in the right place at the right time. That will never happen if you don't turn your radio on.

Never operate under the onus that you don't have what it takes to make a new radio discovery…or re-discovery…as a lowly radio amateur. Allow me to use as a case in point, X and O mode propagation. I would *love* to be able to say that I discovered the phenomenon. In reality, the basics of Extraordinary and Ordinary waves were known nearly 200 years ago, in the field of optics, before "radio" even had a name. Physicists and other professionals knew all about X and O ionospheric propagation over 70 years ago, and it had been discussed in *World Radio* magazine by Bob Brown, NM7M and Carl Luetzelschwab, K9LA, back in the 1990s. I learned about X and O propagation working at Hipas Observatory, some 25 years *after* I had been licensed as a ham…and nearly as long in the field of broadcasting. I can take no credit at all for understanding the phenomenon.

What I *can* take credit for is realizing its importance to normal Amateur Radio. My December 2010 *QST* article, "Gimme and X, Gimme an O" was the first ARRL publication in more than 70 years to mention the phenomenon. In classic Amateur Radio fashion, I was just in the right place at the right time: *Proplab Pro 3* makes it easy to demonstrate my X and O assertions, not that they needed any proof.

Knowing about X and O propagation was, perhaps, the first "discovery" in Amateur Radio for a long time that actually suggested a fundamentally new way of actually doing ham radio…using circularly polarized antennas on HF, a preposterous idea!

Amateur Radio needs more preposterous ideas. The best computer program probably won't suggest any preposterous ideas, but it can certainly help confirm them.

We trust the last chapter, "Unexplored Territory," has wobbled a few new brain follicles. We still have a great opportunity to "advance the state of the radio art" as spelled out in FCC Part 97.1, which delineates our "basis and purpose" as radio amateurs. Nowhere is it potentially more likely for the "Average Joe Ham" to make a real contribution to the radio art than in the understanding and exploitation of some frontier of radio propagation.

Walking the Planck: Why Photons Don't Matter (Very Much)

Inevitably, whenever a discussion of radio propagation ensues among hams and other reasonably astute technical types, the matter (in the linguistic sense, not the physical sense) of *photons* comes up. "Why are we talking about rays and electric waves and magnetic waves and fields and stuff like that, when everyone born after the 19th century knows that radio waves are made of *photons*?" is the typical refrain, spoken or implied.

This is certainly a fair question, which we will address here as concisely and painlessly as possible.

Yes, photons exist, and every principle we discuss in this book can be reframed in a modern quantum-mechanical understanding of nature. With sufficient mathematical gymnastics, we can describe radio behavior in terms of streams of particles (photons) emanating from antennas and other generators. However, as it turns out, it's just not very efficient or helpful to do so. *Classical* electromagnetism can explain about 95% of everything we experience in radio propagation, and *Newtonian* mechanics can usually explain another 4.5%. The remaining 0.5% of what we observe as radio amateurs yields best to "photon thinking."

A Small Fly in a Big Ointment

There's a little number, a *very* little number, known as Planck's Constant that basically shows us how irrelevant photons are at frequencies we normally use for radio communications. (A few of us do experiment with optical communications methods, where this number becomes a bit more meaningful, but this is a somewhat rare exception.)

Let's look at the number:

Planck's constant (h) = $6.62606957 \times 10^{-34}\ m^2\ kg / s$

Even without being a mathematical whiz, we can deduce that anything to the minus 34^{th} power is going to be really small. But what does Planck's constant really mean? When everything is boiled down, it turns out to be a *proportionality* constant, which relates Energy to Frequency by the formula:

$E = hv$

where

E is Energy

h is Planck's constant

v is frequency.

We don't need to derive any actual values to demonstrate that if h is extremely small, v must be extremely large in order for E to be meaningful.

For a seasoned "radio guy" or "radio lady" this concept is not all that intuitive. In radio work, we generally consider *energy* and *frequency* to be two totally independent values. A radio wave has a certain *power* level (energy per unit of time) and a frequency, and never the twain shall meet. We never consider that a 100 W, 10 meter radio signal actually has more energy than a 100 W, 80 meter signal. Planck's constant will tell us that the difference in energy levels between the two frequencies would be so minuscule that it is not measurable by any known instrumentation.

It is only when the wavelengths involved become on the order of the dimensions of an atom that photons become meaningful, or put another way, that Planck's constant becomes a large number relative to other things.

When considering the ionization process due to solar UV radiation, Planck's constant becomes meaningful. The energy difference between long UV and short UV becomes significant, as does the difference between UV and X-ray radiation. X-rays have much larger energy and ionization capability than long UV rays. Photons in this regard become a more useful "medium of exchange" than they do at "normal" radio frequencies.

Bunches

If we now know that the energy (or power) of photons is a function of *frequency*, then how in the world do we "translate"RF power to the photon world? It's quite simple. If the energy level of a photon is a measure of its *frequency,* the *number* of photons tells us how much RF power we have. So a 100 W 80 meter transmitter pumps out more photons per second than a 50 W, 80 meter transmitter. In either case, we're pumping out *bunches* of photons, but the energy level of each photon is small beyond minuscule.

You are certainly welcome to use photons for all your radio work. I just don't envy you working with those kinds of numbers!

Understanding AM

There are two kinds of hams in the world: those who use AM and those who would like to.

Archaic Modulation, Ancient Modulation, Aztec Modulation…call it what you will… there has always been a certain primitive charm to old school amplitude modulation. Long considered the sole domain of crusty old "Boatanchorologists," AM still has a few tricks up its sleeve. Modern hardware (and software) allows the modern ham to explore this mode with little difficulty or danger to life and limb, something that couldn't always be said of the first generation of AM experiments. We'll look at a few interesting nooks and crannies of AM, which will hopefully suggest some ideas for experimentation.

To the best of our knowledge, amplitude modulation was the first method available for applying voice to a radio signal. Because of its sheer simplicity, it's easy to overlook some of the more obscure and fascinating aspects of this time-honored technology. The fundamental method is implicit in the name, amplitude modulation. It is the art of changing the *amplitude* of a radio frequency signal in accordance with the intelligence (or lack thereof) that one wishes to transmit. Some of the earliest methods of achieving this end were painfully and frighteningly simple, such as inserting a carbon microphone in series with the high voltage lead of a spark gap transmitter! A number of early experimenters survived to tell about this alchemy…er…*technology*.

Needless to say, the audio character of such a horrifying device left something to be desired. The "note" of a typical spark gap Morse code transmitter in itself was about as pleasant on the receiving end as a chainsaw cutting through a porcelain sink; the addition of voice modulation to the cacophony only added to the tortured screech. Fortunately for all involved, amplitude modulated spark gap transmitter technology never really caught on in a significant way; it took the advent of an actual *continuous wave* (CW) transmitter to form the foundation for practical voice modulation. We need to be careful not to denigrate the aforementioned Medieval Method of Modulation too much, however, because it did provide a basis of understanding for what was to follow.

Options…of Sorts

When one looks at an unmitigated radio frequency carrier, one realizes that there aren't a whole lot of things you can do to the thing, if applying intelligence is your goal. You can change its strength (amplitude modulation). You can change its frequency (frequency modulation). You can change its direction (directional modulation). (Yes Virginia, there is such an animal). You can change its phase (phase modulation). And you can change its polarization (polarization modulation). Hmm…maybe you do have a few interesting options. But AM came first, and it's lasted the longest. And it's also the easiest to analyze mathematically.

Math Schmath

As it turns out, radio amateurs (and commercial radio folks, too) were producing some pretty decent AM signals without much of a mathematical basis for it. It's not that the appropriate math didn't exist. That famous dead French mathematician Joseph Fourier had pretty well described all the mathematics we needed a couple of centuries before we needed it. But somehow, we managed to muddle through without it, largely through the incredibly precise application of trial and error…mostly error.

In fact, for much of the early history of AM radio, little or nothing was mentioned about *sidebands,* for instance. And it's still true that you can build a pretty decent AM transmitter without knowing diddly-squat about sidebands. But you can do it even better if you *do* know about such things.

What AM Isn't

On the surface it may seem that an easy way to create amplitude modulation is to simply add an audio voltage (**Figure A2.2**) on top of a radio frequency carrier (**Figure A2.1**). Take some radio, take some audio; throw it all in a pot and viola…AM radio. Unfortunately it doesn't quite work this way. If you take a radio frequency carrier and add an audio voltage on top of it, what you end up with is a thing called a radio frequency carrier with an audio voltage added on top of it (**Figure A2.3**).

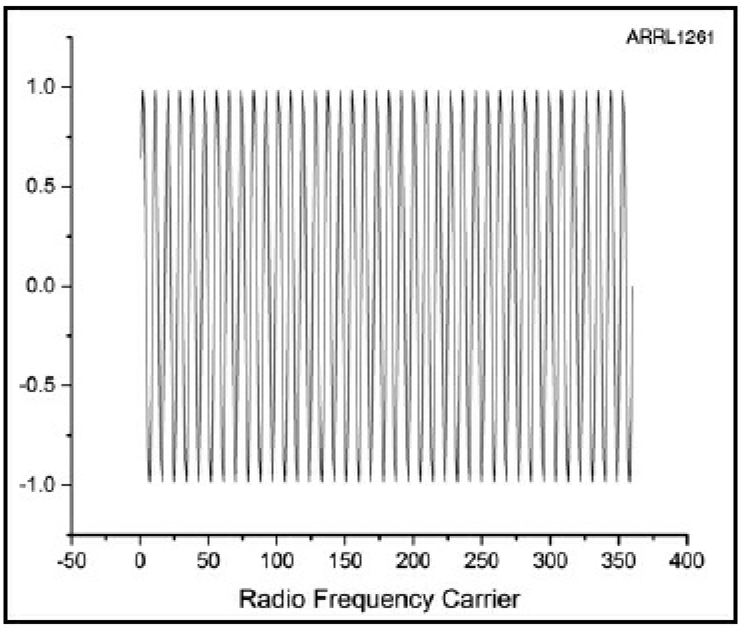


Figure A2.1 — An unmodulated AM carrier is a simple sine wave. It has zero bandwidth. The *envelope* of the carrier is defined by the upper and lower peaks of its amplitude.

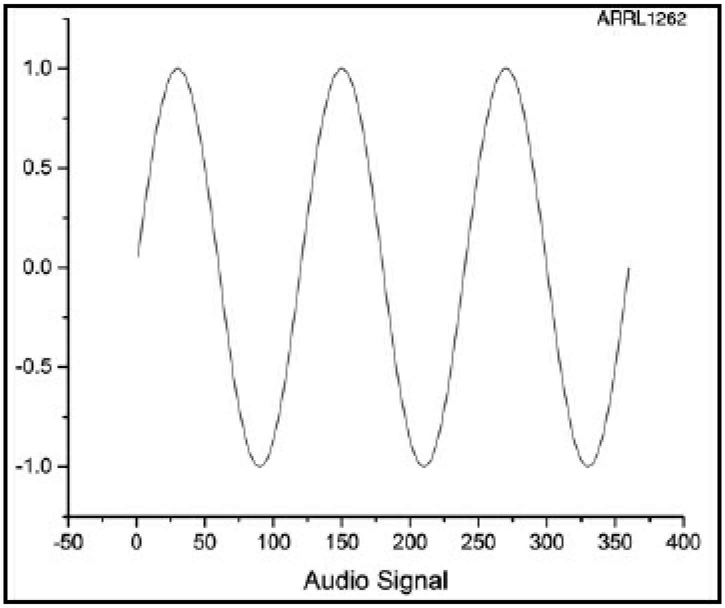


Figure A2.2 — A modulating signal is much lower in frequency than the carrier. For simplicity of analysis, we will use a single audio sine wave; in "real life" the modulating signal is much more complex… but can be analyzed as the summation of a number of sine waves if varying frequency, amplitude, and phase.

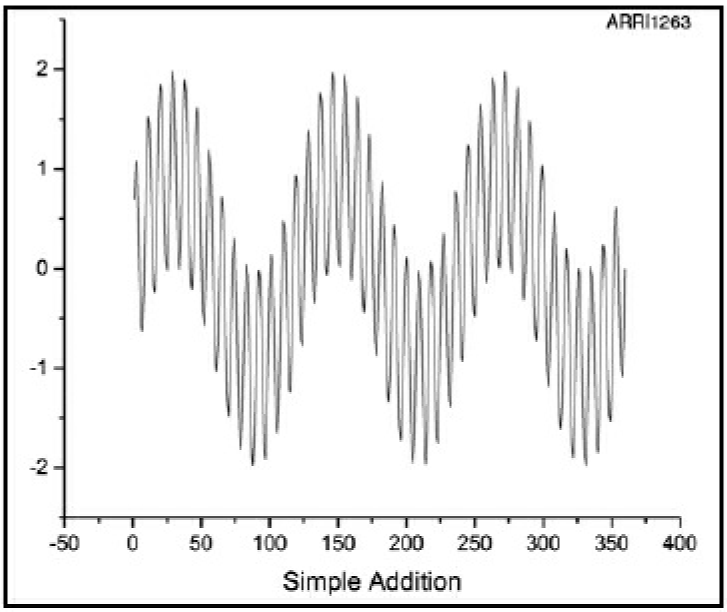


Figure A2.3 — What AM modulation *isn't*. Simply adding an audio signal (in this case a pure sine wave) to a radio frequency carrier simply results in to frequencies occupying the same space. The individual signals retain their complete identities, imparting no information to the radio signal whatsoever.

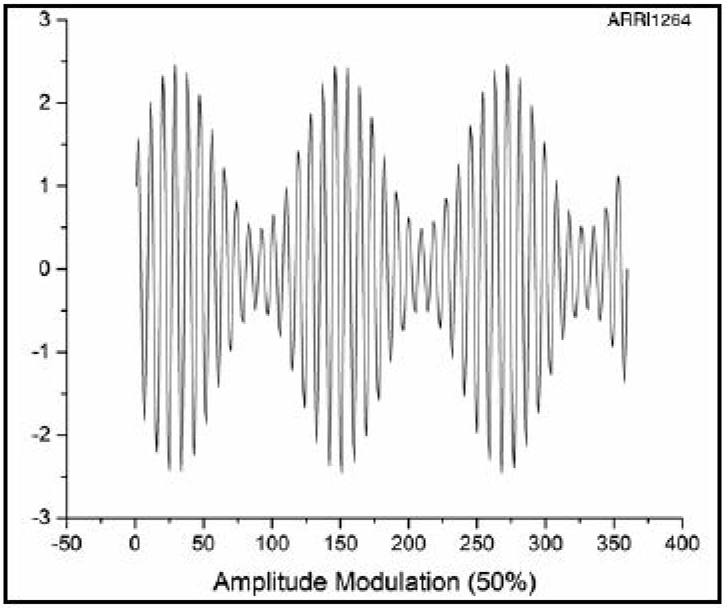


Figure A2.4 — That's more like it! Here we have genuine AM modulation happening, wherein the amplitude of the radio frequency carrier is *modulated* by the audio. The AM process also generates *sidebands*, which are separated above and below the carrier by the audio frequency. These two sidebands, being different in frequency may have considerably different propagation paths, resulting in *selective fading*. This was one of the primary motivations for the development of single sideband, where selective fading is greatly reduced.

Mathematically speaking, the result shown in Figure A2.3 is simply the algebraic *addition* of the two voltages. In a linear system (that is, one that follows Ohm's Law), the principle of *superposition* holds forth. Superposition tells us that each frequency maintains its original identity, with no interaction whatsoever. It's a good thing Superposition exists, or we would live in a very odd universe indeed! Superposition "works" in free space, in wires, and in resistors. The superposition of audio on top of a radio frequency carrier does *nothing* to the radio frequency carrier. The amplitude of said radio frequency carrier remains absolutely untouched, and so we have no information applied whatsoever.

What we *really* want to have happen is shown in **Figure A2.4**. This is the result of algebraic *multiplication* of the two signals. (This is absolutely *not* to be confused with *frequency multiplication*; it is the *amplitudes* of the signals which are multiplied together). This is not to say that frequency multiplication cannot *also* take place at the same time, but that is a secondary effect. The primary result we want is the multiplication of the radio frequency amplitude by the audio frequency amplitude (at any given point in time).

How We Do It

In order to perform the multiplication function, we need some *nonlinear* element in the cauldron. This can be a diode, a Class C amplifier (either vacuum tube or transistor), or combinations thereof. You *cannot* use a linear RF amplifier to create AM, although you can certainly use one to *amplify* an AM signal that's already been created "upstream." A transmitter that performs the modulation process in the final stage is known as "high level" modulation, while a transmitter that performs the modulation upstream and then feeds the results to a linear amplifier is known as "low level" modulation. Most modern rigs use low level modulation for AM, whereas the aforementioned AM spark gap transmitter clearly falls into the high level modulation category.

Heavy Iron

The most straightforward means of modulating a Class C amplifier is with a *modulation transformer*. In this configuration, the plate voltage of the final RF amplifier is passed through the secondary of a heavy iron transformer. The primary of the transformer is driven by the audio, typically via a push-pull class B amplifier, but on occasion using a single ended Class A amplifier. In either case, a significant amount of audio power is necessary, thus requiring a fairly substantial transformer. To fully modulate (we'll talk about percentage of modulation shortly) an AM transmitter requires that you supply 50% of the RF amplifier's dc input power in the form of audio. For example, if you have a 100 W output RF amplifier (unmodulated), and the amplifier is 75% efficient, the dc input power to that stage will be 133 W. This means you need to supply at least 66.5 W of clean audio power into the modulation transformer. Now, it's not always necessary to fully modulate (100%) an AM transmitter (**Figure A2.5**). Many times there are advantages to "undermodulating" a transmitter (Figure A2.4 again). Many military and aviation transmitters use fairly low levels of modulation. But for working DX, you want to be able to modulate as "deeply" as possible. As we will see, the peak power output increases as the *square* of the modulating (or *envelope*) voltage.

Keep in mind that while there is no theoretical maximum *positive* level of modulation (maximum peak carrier amplitude), there is a theoretical maximum *negative* modulation, where the carrier is completely *pinched off* at the middle as in Figure A2.5. You can't transmit *less* than zero power… although attempting to do so will create horrendous distortion!

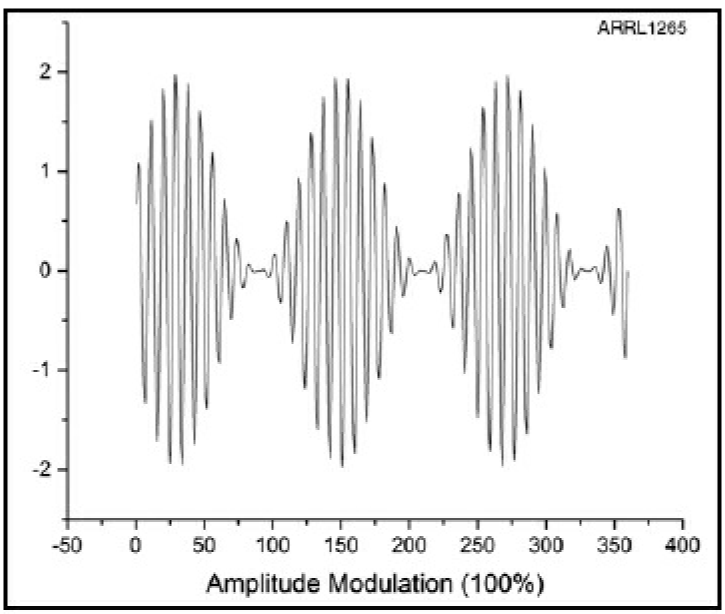


Figure A2.5 — Here we also have AM modulation, but with greater depth. This is the maximum modulation one can achieve with a sine wave audio signal as the carrier level is "pinched off" to zero during the negative modulation peaks.

Square Law Stuff

In our discussion so far, we have shown only the *voltages* involved in the AM modulation process. However, if we're transmitting into an actual load of some sort (say 50 ohms) the actual power at any time is described by E^2/R. So, doubling the peak amplitude is four times the RF power! A fully modulated AM signal therefore has 4 times the peak envelope power of the unmodulated carrier. So you can see that any effective high power AM transmitter has to have some reserve power capacity all around.

Not So Heavy Iron

Because of the size and expense of "normal" high powered AM transmitters, clever hams looked for ways of circumventing the modulation transformer. As it turns out, modulating the screen grid of a tetrode or a pentode tube operating in Class C results in fairly high levels of modulation with minimal applied audio power. You could also modulate the control grid, but with a lot more distortion in the final product. This was sometimes known as *efficiency* modulation, but more commonly referred to as *deficiency* modulation, because of the generally poor audio quality. (Despite the size and weight, the classic heavy iron transmitter is capable of the cleanest signal, with regard to audio distortion. This is why it was long the preferred method of AM broadcasting, as well.)

Broadcasters themselves weren't averse to experimentation, either. Several interesting (and proprietary) systems came out of the broadcast biz, things like Doherty modulation, the RCA Ampliphase system (ages ahead of its time), and various forms of semi-digital Pulse modulation. (I doubt there are any hams running a surplus Ampliphase transmitter, but one never knows!)

Pushing the Envelope

As soon as hams discovered the oscilloscope, they discovered that audio voltage waveforms are not symmetrical….transmitting the human voice is not the same as transmitting audio tones. In fact, they can be *very* asymmetrical…often by a factor of 6:1 or more. In itself, this is not a problem, but if you wanted to really modulate heavily, you could overshoot on the *negative* modulation peaks *long* before you approached full modulation on the "high" side. Again, clever hams discovered that if you could *intentionally* "flip" the modulation polarity around so that the larger peaks were always positive, you could get a lot more peak envelope power out. This led to *supermodulation*, which was probably the next best thing to single sideband. Probably the most interesting (and controversial) part of this was that, according to the FCC rules at the time, your power limit was determined by the dc *input* power to the final stage (2000 W). This did *not* account for modulation whatsoever! What this meant was that, in the 1950s, one could *perfectly legally* transmit up to around *four times* the present day PEP single sideband power output limits! (This was using only the *natural* asymmetry of the typical male voice; with some actual processing, this power figure could be even higher!) Alas, with the current FCC rules specifying *output* power, one can no longer use the "AM Cheat." How sad.

Variations on a Theme

I have a prized, pretty rare boat anchor, a Central Electronics 100V. It was one of the first commercially available single sideband transmitters, but folks weren't quite ready in 1959 to abandon their familiar AM altogether. This gem has every imaginable variety of AM: Double sideband with carrier, double sideband suppressed carrier, upper sideband with carrier, lower sideband with carrier, upper sideband suppressed carrier, lower sideband suppressed carrier, and FM thrown in for good measure. (It also has some really ahead-of-the-curve features for its time, such as broadband tuning, but that's a topic for another time). The point is, there was a semi rocky road toward the widespread adoption of single sideband, and a few of the intermediate steps are worth noting.

A standard AM signal occupies a bandwidth equal to twice the highest audio frequency. The voice energy is contained in two (redundant, by the way) sidebands. Well, not exactly redundant, but mirror images. But the effect is the same. You knew we'd eventually get around to sidebands, didn't you? Well, not to fear. This is actually some pretty neat stuff. If you have a "full AM" transmitter and want to pass audio from, say, 300 to 3000 Hz, your occupied bandwidth will be 6 kHz as shown in Figure A2.4. (You will also have a 600 Hz "hole" in the center of your occupied spectrum, as well, which is often forgotten about by many AM operators). There's no reason you couldn't fit a few CW stations inside that hole…but for some reason, this was never normal practice, even during the heyday of AM.

At any rate, you can improve overall system effectiveness, by eliminating one sideband, or the carrier, or both. (One April *QST* article from the era suggested "suppressed sideband, full carrier" as a viable option as well). If you don't get the joke, stick with us a moment.

One of the inevitable byproducts of the AM modulation process is the generation of *sidebands.* Mssr. Fourier's wonderful equations explain this (as if one can ever consider an equation an "explanation" for anything). But as it turns out, sidebands are more than mere appendages; indeed they carry *all* the intelligence (or stupidity) of an AM signal! You cannot have AM without sidebands, and without sidebands, you have no AM!

Which brings us to a chicken and egg paradox of sorts. Are sidebands really the product of the AM modulation process, or do sidebands *create* AM? As it turns out, both are true. One way you can create a "full AM" signal is by means of three signal generators, one tuned to the carrier frequency, one tuned to the lower sideband frequency, and one tuned to the upper sideband frequency. You can feed these three signals into a *linear* combiner (such as three resistors) and lo, a genuine AM signal emerges. (You do have to appropriately set the relative phase of the three signals, by the way). Nobody actually generates AM this way (at least nobody in their right mind), but it's doable.

A Foray into Fourier

Nobody hates math more than yours truly, so I like to use "thought experiments" to convey some self-evident principles that are a lot easier to grasp than things like Fourier Analysis. Fourier, by the way, narrowly escaped beheading during the French Revolution, supposedly due to his politics. Personally, I think it was because of one of his math classes, but that's neither here nor there.

Imagine a radio frequency sine wave of one frequency, say 1 MHz. As long as it's a pure sine wave, it has one and only one frequency. Now, let's take that sine wave and gradually decrease its amplitude. Most semi-astute radio folks would say, "Oh, that's a sine wave with a gradually decreasing amplitude." But is it, really? As it turns out, you *cannot* change the amplitude of a sine wave without changing its *shape* somewhere along the line. Each cycle will have a slightly different shape than the one before. The shape and size are inextricably linked. So, what we have is no longer a sine wave. In the process of changing the size of the wave, we also incrementally add other frequencies, by virtue of also changing its shape. The more quickly we change the shape, the greater the difference frequency between the sidebands and the original sine wave.

Now, as bizarre as some of this may seem, it is what makes it possible to manipulate sidebands independently, and thus create all these wonderful newfangled modes like single sideband, suppressed carrier. The math may be complicated, but the hardware is pretty simple….everything boils down to using the right filter.

Keeping Balanced

One of the most common ways of generating any kind of AM signal these days is by means of the balanced mixer (most commonly the doubly-balanced sort). A DBM is really nothing more than a precision *multiplier*. Internally, it consists of four diodes in a ring configuration (hence its alternate name *ring modulator*) and a couple of center tapped transformers. A modern DBM has three ports: RF input, IF output, and local oscillator (LO). These building blocks can be used in a wide variety of applications, and as a modulator, it's very versatile. IF you apply an RF carrier to the RF port, and audio to the LO port, you get *double sideband* suppressed carrier out the IF port. You can also unbalance the DBM by means of a small dc voltage superimposed on the audio going into the LO port; this gives you AM with an adjustable carrier level. AM generated in this manner is extremely clean, and can easily be followed by a chain of linear amplifiers to yield a world class high power AM signal. You can control the amount of carrier power by the dc offset voltage, and the modulation depth by the audio voltage. Pretty slick.

Where AM Shines

Since the overall efficiency of a single sideband suppressed carrier system is vastly greater than that of an AM system, is there any reason, other than nostalgia, for the radio amateur to be interested in AM? Absolutely.

One of the problems with a single sideband suppressed carrier link is that the frequency of any detected audio signal is strongly dependent on the frequency accuracy of a number of components, both in the transmitter and in the receiver. Everything must be perfectly matched to "put everything back together" at the receiving end. Now, with the advent of things like precision synthesizers, this is nowhere near the problem it once was, there are still applications where *absolute* frequency stability is a must.

Imagine you have an HF telemetry setup, where you need to know *exactly* the frequency of an audio signal at the transmitting end. Sure, you could connect the audio frequency into the mic input of a modern synthesized SSB transmitter, and with a well matched receiver, extract that audio frequency at the distant receiver. This is assuming every local oscillator in the system is right on the money.

With a full AM system, there is absolutely no error. The "reference oscillator" (the carrier) goes along for the ride! No matter how many frequency conversions the signal may encounter along the way, the demodulated audio frequency will remain unchanged. (By the way, the 440 Hz, 500 Hz, and 600 Hz tones that WWV sends can be precisely received with an inexpensive receiver with atomic accuracy, because WWV transmits single sideband *with* carrier).

Several navigational systems use AM as well, for the same reason.

AM really helps for some digital modes too. Digital modes can consist of combinations of phase shift keying and amplitude shift keying, (QAM). Detecting frequency shift is pretty easy with a single sideband system, but what about amplitude shift…especially if you have several different *values* of amplitude shift to contend with? It can be a real problem if you don't have a reference carrier level. With full AM, you always have a carrier level to refer to, even with fading. (*Selective* fading can still be a problem with AM, but generally this is much slower than the symbol rate, so it still helps to have that old carrier along for the ride!)

Setting Up Your Newfangled AM Rig

If you're a recently minted ham, you probably stumbled across the AM mode on your HF rig by accident. Most old time Boatanchorologists don't need much help getting a decent AM signal out of their heavy iron, but if you're using a rig designed for SSB, there are things to keep in mind. If you anticipate using AM, set the carrier output level for *no more* than 25% of the peak output capability. If you have a nominal 100 W rig, set the unmodulated carrier for no more than 25 w. This may seem a bit wimpy, but remember our square law discussion; you will indeed reach your full 100 W on voice peaks. If you have an unsymmetrical voice (the only way to really check this is with an oscilloscope), you may be able to reduce the carrier level even further, while still reaching 100% on the positive peaks.

Isn't this Cheating?

Some purists consider it a travesty of some sort to use a state-of-the-art rig for AM; it's sort of like having an outboard motor on a Viking ship, according to their thinking. In the long term, this is probably true. But if you're not sure you want to invest in the heavy iron necessary to operate "Real AM," using your 21st century rig is perfectly acceptable for getting your feet wet, and learning how AM is done. For a true feel of AM operation, read my November 1993 *QST* article, "Solder to Talk." Be forewarned, AM operation is addictive, and once you let a boat anchor into your shack, more are certain to follow.

Diversifying Your Options

When performing radio communications in any environment other than free space (and this covers nearly all practical amateur radio communications), there is always a high likelihood of *multipath* conditions. This is true for frequencies ranging from medium wave clear up to microwave. This can be a mixed blessing. On the plus side of the ledger, it means there is often more than one way to get from Transmitter A to Receiver B, and vice versa. The most celebrated case is on HF, where you often have the option of “going long path,” or “going short path.” An unexpected long-path signal is generally cause for excitement, rather than an annoyance to contend with (except on some extremely rare occasions).

On the liability side, two simultaneous radio paths between two points can often result in destructive interference. A medium wave groundwave signal combined with a skywave signal can result in effects ranging from moderate distortion to total signal cancellation.

In addition to multipath, we can also have rapid, hard to deal with variations occurring along a single path. Radio signals can significantly change skip distance, polarization, and angle of arrival over surprisingly small time scales.

A number of so-called *diversity* methods can address and reduce most of the vagaries of radio propagation, including multipath and fading. While these methods are commonly employed in commercial and military systems, they are seldom taken advantage of by radio amateurs.

Spacing Out

Perhaps the most familiar form of diversity reception is *spatial* diversity. This method was commonly used to deal with *selective fading,* which is caused by slightly different propagation paths taken by different frequencies due to ionospheric *dispersion.* To the uninitiated, it can seem a bit surprising that a few hundred hertz difference between radio signals could result in even *measurable*, much less *harmful* selective fading. But such is the case indeed. The difference in MARK and SPACE frequencies can be dispersive enough to cause complete annihilation of RTTY signals. For a relatively broadbanded signal like AM phone, the distortion caused by selective fading can be significant, and this was one of the strong incentives for the development of single sideband and, more recently, synchronous AM.

The simplest method of reducing selective fading is by installing widely spaced receiving antennas, where the likelihood of both antennas undergoing the same fading conditions simultaneously is greatly reduced.

A typical diversity system consists of two antennas, two separate, identical receivers, and a *post detection* combining system (usually at the product detector level). Some military receivers such as the venerable R-390s had built-in provision for diversity reception.

This isn’t always a practical option for the budget-strapped amateur, who may have a difficult time scratching together two receivers, much less two *matching* receivers.

Raising Your IQ

There are a number of applications in radio where having two or more identical receivers would be nice. The once-disparaged *direct conversion (DC) receiver* has found an irreplaceable niche in *radio science*. Modern components make possible the construction of extremely high performance and inexpensive DC receivers, but even more importantly, DC receiving methods allow the construction of many *identical* receivers. In addition, it is very easy to deploy DC receivers in *phase quadrature,* providing a whole new dimension in flexibility, not only for diversity methods, but also in-depth radio studies. An "I/Q" receiver (In phase and Quadrature), simply uses two detectors, shifted in phase by 90 degrees.

The block diagram in **Figure A3.1** shows a representative multiple IQ receiver system, which can be expanded to as many receivers as desired. The only critical component in the system is the local oscillator, which needs to be clean and stable. The good news is, you only need one of them, no matter how many receivers you have!

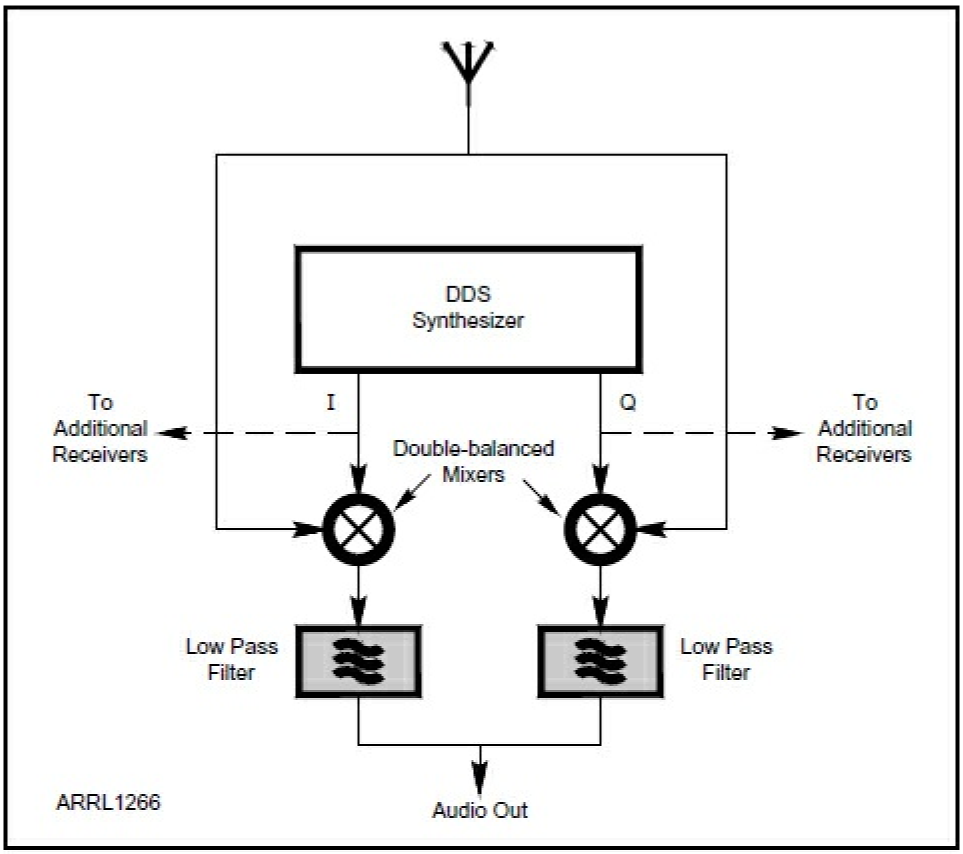


Figure A3.1 — An I/Q direct conversion receiver is at the heart of a diversity reception system. Multiple receivers can be driven from a single oscillator, in this case a direct digital synthesizer.

For the dedicated experimenter, we cannot recommend the Novatech quadrature DDS 8m synthesizer too highly: **www.novatechsales.com/PDF_files/dds8m.pdf**. This module provides two very clean quadrature signals (sine and cosine) using direct digital synthesis (DDS). You can program this to operate from dc to 100 MHz, as well as perform many sweep and modulation functions.

The price of $695 may seem steep, but again, you only need one of these gems to form the core of a world class scientific receiver array. You can even build a full-fledged spectrum analyzer around the unit!

More Polarizing Thoughts

The same IQ receiver system can be used to obtain polarization diversity, such as when using a separate vertical and horizontal dipole. Or perhaps, more revealingly, a clockwise CPOL antenna and a counterclockwise CPOL antenna. For linearly polarized signals that slowly or randomly change polarization, this can greatly reduce fading. For known CPOL signals (X and O waves), this can be used to quickly determine *which* of these waves you're dealing with.

The Plot Thickens

Although we refer to the detected output of our DC receivers as *audio*, in many cases this is nothing more than a dc voltage. This will often be the case when you're doing propagation experiments with WWV or other beacon stations. One of the fascinating things you can do with any IQ receiver is to use an oscilloscope in the XY mode to plot the detected voltage. Connect the I output to the X channel of the scope and the Q output to the Y channel of the scope. If you use, say, a vertical/horizontal diversity antenna, you will see the dot move to the axis where the greatest signal is. You will be surprised how much this moves around under even the most supposedly stable conditions!

Doppling for Fun and Profit

One of the neatest things that an I/Q receiver shows is Doppler shift. You don't need more than one receiver or antenna to show this. A single IQ receiver connected to an XY oscilloscope will show you just how fast the ionosphere can move around. If a signal is moving lower in frequency, you'll see a dot move around in a circle one direction; if it's moving higher in frequency, it will run around the screen in the other direction. During nightfall, the ionosphere can retreat by as much as 700 feet per second, which can cause the trace to spin around at quite a merry rate!

Doppler Diversity

As much fun as it is to watch ionospheric Doppler in action, it doesn't do much for the enhancement of communications. As any satellite operator will tell you, Doppler shift can be a real pain. However, the effects of Doppler shift on HF are a bit more subtle…and unlike a low orbiting satellite which will *always* have Doppler, the ionosphere only exhibits significant Doppler during sunrise or sundown, as it's either advancing or retreating. During the daytime or nighttime, there may be localized movements and waves in the ionosphere, but on the average it's fairly stable.

All this being said, if you were to desire to create a Doppler cancelling diversity scheme, it would require two different frequency receivers, one slightly above and one slightly below the radio frequency of the signal of interest. For a pair of DC receivers, this would suggest a frequency separation of a few tens of Hertz…the typical Doppler shift frequency of a receding ionosphere in the mid-HF frequency region. It might only be really useful during the fifteen minutes or so when this activity is in full force, but it could reveal some interesting physics.

The Goniometer is Not Gone

One of the most interesting radio devices of days past is the *radiogoniometer*, a radio direction finding device that did not need to physically steer the antenna. It consisted of two loop antennas at right angles. The outputs of these antennas fed a pair of orthogonal coils with a rotatable *tickler* coil inside of them. By rotating the tickler coil for maximum signal, you could deduce which direction the signal was arriving from…minus the ever-confounding ambiguity that a "normal" loop antenna gave you. It was an extremely clever and effective device, but did require extreme care and a lot of shielding around the tickler assembly.

A 21st century incarnation of the radiogoniometer can be implemented using a pair of crossed dipoles (or loops if so desired) and a *single* I/Q receiver. One antenna feeds the I channel, and the other feeds the Q channel. The outputs of the receiver are fed, again, to an XY oscilloscope. The direction of arrival of a signal can be determined by the location of the spot on the oscilloscope. But here's the real neat part. Because of the 90 degree phase difference between the channels, the 180 degree *directional ambiguity* of either antenna is eliminated *unless the signal is arriving PRECISELY from a north-south or an east-west direction*! A signal arriving from any other direction can result in *only* one "quadrant" being activated on the oscilloscope.

Now, unless your receiver is actually phase locked to the incoming signal, there will usually be a few hertz difference between your local oscillator and the incoming signal. This will cause the display to not be a fixed dot, but rather a dot that moves slowly in an ellipse…the orientation of the ellipse being representative of the direction of arrival.

Roll Your Own

As you can see, the multi-IQ receiver setup can give you nearly limitless possibilities for various types of diversity reception. There are undoubtedly some forms of diversity reception we haven't thought of. Again, the DC receiver is a cheap, versatile means of doing some real radio science. We encourage you to explore this fascinating technology, and do feel free to share your ideas and findings.

Locking in to Some Real Science

It's exciting to see the renewed interest in weak and very weak signal modes among amateurs, largely propelled by the advent of the "JT" modes. The well-deserved buzz about this family of new digital modes speaks well of the health of Amateur Radio experimentation. These weak signal modes allow even the most impoverished ham access to a realm of Amateur Radio once reserved for the most exotic research labs.

Nevertheless, before we become too enamored with any technology just because it's *digital,* we need to recognize that even the most astute bit-twiddler has to work within the rules of physics, especially that most unforgiving rule of physics, the NFL (no free lunch) principle. In weak signal work, NFL manifests itself in the law that says the weaker the signal, the longer it takes to detect it. JT65 is extremely sensitive, but it's also extremely slow. At least in *this* universe, there's no way of escaping this dichotomy. While digital and analog methods both approach the NFL problem from very different angles, they still end up having to solve the same problem.

This brings us to the relatively ancient methodology of the *lock-in amplifier*, which in certain cases is capable of detecting coherent radio signals 60 dB or more below the noise floor. Lock-in methods are used in everything from seismology to materials testing, but hold a valuable place in weak signal radio work, as well. Like our now-familiar weak signal digital modes, the Lock-in amplifier is subject to NFL; it takes a long time to do its magic.

The happy converse of the NFL equation however is this: *If you have enough time, you can detect anything. (*Provided, of course, the thing you're trying to detect sticks around long enough for the detection process to work.)

The Hardware

The lock-in amplifier is essentially the same as the IQ receiver, which outputs in-phase (I) and quadrature (Q) channels for use with digital signal processing (DSP) software. (See "Raising Your IQ" in Appendix 3.) The primary difference is the post detection audio filter. While normally we characterize the filter in terms of bandwidth, in the lock-in amplifier, we speak of *periods* or *integration times.* In ELF and ULF work, we may be dealing with radio frequencies on the order of millhertz. The period of a millihertz radio signal is in the thousands of seconds. Depending on how many cycles we want to *integrate,* we could have integration times on the order of *hours.* So, basically, our receiver's bandpass filters become integrators with really long time constants.

Weak signal work, especially down in the 500 kHz experimental band, primarily consists of very slow CW, sometimes known as QRSs, or QRSss, or QRSsss. There are no firm definitions of these speeds, but in the weakest signal cases, we're dealing with fractions of a character per minute.

Now, the real tricky part of the lock-in amplifier is that the local oscillator has to be phase coherent with the transmitted signal you're looking for. This is, of course, not always practical…or even possible. However, with the widespread availability of GPS, it's a lot *more* possible than it was in the past. It's not too difficult to build a *disciplined* oscillator at the distant transmitter that is phase locked to a GPS satellite. The transmitter's frequency can then be known to extreme accuracy. The local oscillator in the lock-in amplifier can likewise be disciplined by the same GPS. Short-term errors in the GPS frequency, due to Doppler shift or other factors, are swamped out by the very long integration times, while long-term phase lock is maintained.

Now, in the case of QRSs, you have to choose your integration times carefully. If they are too long, you'll end up filling in the gaps between dashes and dots even at the slowest speeds. So the integration time needs to match the symbol rate to some degree. There's a bit of an art form to this, so nothing replaces a little experience.

Hybrids

Commercial lock-in amplifiers can be a bit on the pricey side. At the core they are simple analog instruments, but most of them are hybrids that incorporate precision DDS (direct digital synthesis) to generate the local oscillator signals (both sine and cosine), as well as some elegant DSP in the output filters. In addition, the front end devices are often extremely low-noise semiconductors, which can be fairly expensive too. A lot of this cost can be reduced by "rolling your own" DDS or DSP with Arduino or similar technology, and we expect to see the price of lock-in amps fall dramatically as more hams apply these tools. In the meantime, lower-tech, all-analog lock-in amps can be built by anyone who is capable of building a direct conversion receiver.

Final Notes

Analog technology will always be around. This is because radio itself is primarily an analog phenomenon. It never hurts to have a few alternative techniques in your toolbox. It's always best to fit the technology to the task.

The Invisible Journey!

There are countless ways for radio signals to travel from transmitter to receiver, and understanding how radio waves interact with their environment is an important factor in successful radio communications. While radio amateurs can maximize station performance and reliability with the right equipment, knowledge, and skill, we cannot control propagation. Through scientific exploration and experimentation, we can improve our understanding of propagation and how it affects radio signals.

Propagation and Radio Science presents a comprehensive overview of one of the most fascinating and rewarding aspects of Amateur Radio. Author Eric Nichols, KL7AJ, uses his lively, engaging approach to present the complex subject of radio propagation in simple, easy-to-understand terms. This book covers topics ranging from theoretical exploration to practical application. It explains the phenomena we observe on the amateur bands and invites you to embark on the journey through the still-unknown radio propagation universe.

Chapters include:

- Matters About Matter
- The Optical Factor
- Polarization, Gain, and Other Antenna Matters
- The "Reflection" Process
- The Ground Wave
- Demystifying the Ionosphere
- The Anomalous Ionosphere
- Magnetic Personality
- Instrumentation and Interpretation
- Free Electron Propagation
- Neutral Propagation
- Cheaper Than Dirt
- Diversity Methods
- WWV and Channel Probes
- Software and Other Tools
- Keeping Up with Kepler and Friends
- Your Friend the S Meter
- Loads of Modes
- Sea Shanty
- NVIS Modes and Methods
- Unexplored Territory

About the Amateur Radio Service

Amateur (Ham) Radio provides the broadest and most powerful wireless communications capability available to any private citizen anywhere in the world. The principles of this federally licensed radio service include public service, radio experimentation, training, and international goodwill. ARRL is the national membership association for Amateur Radio operators. ARRL has books, software, online courses, and other resources for licensing, operating, and education.

Published by:
ARRL The national association for AMATEUR RADIO®
225 Main Street, Newington, CT 06111-1494 USA
www.arrl.org