COMMUNICATIONS
QUARTERLY

THE JOURNAL OF
COMMUNICATIONS
TECHNOLOGY

Summer 1999

$9.95

COMMUNICATIONS QUARTERLY

THE JOURNAL OF
COMMUNICATIONS
TECHNOLOGY

CONTENTS

Cutaway detail of brass
rod termination with
nuts and lock
washer on
either side of
U-stock

3/4 x 1/2" channel (

Cutaway detail of brass ro
termination with nut an
lockwasher on
either side o
tubing

3/4" Al. tube   U-stoc

page 25

page 74

On the Cover: Many hams have taken to sharing memories of DXpeditions
and pictures of their stations via computer and the Internet. Here's Bryan
Bergeron, NU1N's vision of how it all works.

A little radio gem in New Hampshire

Traveling in the New England area this summer? Tired of hanging around the same old tourist traps? Have I got a deal for you!

Nestled in historic district of Bedford, New Hampshire, is a small clapboard building that bears the legend "Marconi Museum." In earlier incarnations, the structure was a school, courthouse, town office, and police department. But step inside its double doors and you'll find an *incredible* collection of early radio equipment!

A special investment

The Guglielmo Marconi Foundation, U.S.A., and its accompanying U.S. National Marconi Museum, were founded in 1995 to "promulgate the name of Marconi, 'Father of Wireless,' who 100 years before, demonstrated the transmission and reception of Hertzian waves as a new medium of communication." The Foundation's Chairman, and your host, is Ray Minichiello, P.E., W1BC.

Now, Ray has a special investment in the Foundation and its museum. You see, Ray's family and Marconi's family were good friends when Ray was growing up. His father worked with Guglielmo Marconi at the General Electric plant in Schenectady, New York, and Ray was exposed to all the excitement that took place in the early days of radio. It's no surprise that radio plays a big part in his life.

Ray owns most of the equipment displayed inside this 14,000-square-foot building. (For years it was stored in his mother's house and in his own garage until it found a home in Bedford.) As you tour the museum, you'll see early Marconi wireless equipment* and a progression of radios up through the cellular telephone. There are beautifully preserved specimens in handcrafted, furniture-quality cases with the instructions pasted to the back of the box— no instruction manuals to lose. Stroll along through the rooms and you'll encounter familiar rigs, like ever-popular Hallicrafters, early broadcast microphones, and *enormous* tubes. I know Peter Bertini, K1ZJH, our Senior Technical Editor would drool over the Cathedral and Tombstone radios on display. And that's not all! There are exhibits

*The Marconi Spark Transmitter that is your first stop on the tour was recovered by Ray from a garage in Switzerland after an intensive world-wide search. It is possible that this is the only one still in existence.

of early medical RF therapy devices, test equipment, and even a radio that resembles some sort of primitive space heater.

Other highlights

Be sure to check out the museum's amateur radio station. The antenna is located in the attic, in keeping with the historic nature of the town. You'll also want to browse through The John Frey Technical Library, which contains thousands of radio communications periodicals, some dating back to 1920. The library's shelves hold hundreds of engineering, text, and reference books. I even saw copies of *Communications Quarterly* on the shelves!

The facility has a restoration room for repair of vintage radios, a machine shop, and a facility room for meetings of electronics-oriented organizations. There are also plans to build a 100-watt FM broadcast station in the educational section of the FM band, with a programming emphasis on world scientific news. Because the museum is education-oriented, The Marconi Legacy Fund has been established to provide scholarships to students interested in the pursuit of studies related to wireless communications.

To date, hundreds of items have been collected for the museum, but additional donations of vintage and modern electronic equipment are welcome. As the Marconi Foundation is a non-profit corporation, all donations are acknowledged for personal and tax records.

More to be done

There's much more to be done at the U.S. National Marconi Museum. Ray has many more displays in mind for the future. Lots of equipment awaits the chance to shine in the upstairs galleries. Still this little gem of radio history is well worth the visit.

For those who would like to learn more about The Guglielmo Marconi Foundation, U.S.A., Inc., the address is 18 North Amherst Road, Bedford, New Hampshire 03110. The telephone number is (603) 472-3622. Information is available via e-mail at: <info@marconiusa.org> and don't forget to check their website at <www.marconiusa.org>.

Terry Littlefield, KA1STC
Editor

PRODUCT INFORMATION

Catalog/Website Feature Molex Products

Molex offers two ways to access the latest information on its products. One is the newly published Molex 990 full-line catalog; the other is via their electronic catalog at website <www.molex.com>.

The new 1,598-page catalog contains product listings and incorporates suggestions from catalog users. Many sections have been reorganized by product application for easier selection and ordering.

To request a print copy of the new Molex 990 full-line catalog, contact Molex Incorporated, 2222 Wellington Court, Lisle, Illinois 60532 or call 1-800-78-MOLEX. Molex's electronic product catalog can also be accessed via the company's website: <http://www.molex.com>.

Electronic Devices Catalog Available

Electronic Devices, Inc., offers a catalog featuring low-cost economy-grade bridges. Voltages from 50 to 1000 are available in standard case sizes in leaded, quick-connect, and surface-mount terminations.

For a copy of the CB98 Catalog, write to Electron Devices, Inc., 21 Gray Oaks Avenue, Yonkers, New York 10710, call 914-965-4400, or visit their website at <www.edidiodes.com>.

One-Piece Cable Entry Boots

Andrew Corporation has a one-piece cable entry boot with a self-aligning seam for simple installation and weather resistance of equipment shelter feed-through plates. The boots are available for 4- or 5-inch entry plate holes. Cushions accommodate HELIAX® and other coaxial cables from 1/4 to 3 inches.

The self-aligning seam fits snuggly against the entry plate. Installation is completed by clamping with stainless steel hose clamps.

The boots and cable cushions are made from black EPDM rubber. They have been tested to and meet MIL-STD specifications for thermal shock, humidity and moisture resistance, icing, and vibration.

For more information, request packet 440 (bulletin number 10309) from the Andrew Customer Support Center at 1-800-225-1479.

HP 1999 Catalog of Test System and VXI Products

The 1999 Catalog of Test System and VXI Products is now available from Hewlett Packard Company. The free catalog includes detailed information on the HP's test-system components, VXI products, and support services. It also contains product descriptions, summary application specifications, and photographs on an expanded portfolio of test products including new mainframes, VXI modules, M-Modules, and high-speed interfaces.

Interested parties can order the HP 1999 Catalog of Test System and VXI Products via the World Wide Web at <www.tmo.hp.com/tmo/literature/English/index.html> or by calling 1-800-452-4844, ext. 6494 in the U.S.

Low-Noise Receiver Preamps

Hamtronics' new LNY-( ) series of preamps, a replacement for their LNW-( ) series, offers low-cost preamps designed to mount in the cabinet next to the receiver board. (When ordering, enter the band or frequency you desire in MHz inside the parentheses.) It uses a new low-noise MOSFET optimized for best performance at VHF and UHF frequencies. The FET has built-in diode protection and low feedback capacitance. Surface mount technology is used to obtain minimum noise figure and best stability.

The LNY-( ) preamps have RF-style RCA jacks for coax connections. The pc board is only 1 x 2-1/2 inches, and the unit operates on +12 to +15 volts DC at 10 mA. The pc board also has a provision to add an RF choke to allow the B+ to be fed up the coax when necessary so the preamp can be installed at the antenna without extra power wiring.

Models are available for all popular bands from 28 to 470 MHz. Gain ranges from 18 to 26 dB, and noise figure ranges from 0.6 to 0.8 dB, depending on frequency.

For more details, you can view the entire catalog at Hamtronics' website, <http://www.hamtronics.com>. For a paper catalog, write to Hamtronics, Inc., 65 Moul Road, Hilton, New York 14468-9535 or phone (716) 392-9430.

═══TECHNICAL CONVERSATIONS═══

Using the standard NEC code terminology

Dear Editor:

In regard to Mr. Bertini's reply ("Technical Conversations," Spring 1999), it might be simpler if all concerned used the standard National Electrical Code terminology to discuss what can be a rather confusing subject.

Article 100 of the 1996 edition identifies an Equipment GROUNDING conductor as a conductor that connects the equipment frame to the building frame or service entrance ground. Article 210-5 requires this conductor to have green (or green with yellow tracer) insulation, or to be bare. It normally carries no current except enough to trip the protective device in the event of a phase to ground short.

Article 100 identifies the GROUNDED conductor (commonly called the neutral) only as a conductor that is grounded. Article 210-5 requires it to be covered with white or gray insulation. It connects one pole at the equipment to the matching pole at the service entrance. This leg is defined as the current return. It must be connected to the service entrance ground at that point. It normally carries full load unbalanced return current.

The service entrance is the *only place* where these two conductors are permitted to be tied together. Some very specific exceptions exist but do not usually apply to modern residences. Mr. Bertini is quite correct in his concern about unexpected chassis potential in case of a fault with the proposed circuit. For this reason, buildings are not wired with a GROUNDED conductor (neutral) connected to any chassis ground. All equipment should follow the same practice.

With respect to Mr. Bertini's other question, if the equipment were to be connected to its power source by a plug and socket, a 4-pin Series 14 device could be used as shown in NEC Figure 2.10-7 or -8. It is rated for 125/250 VAC plus GND. Various models in this style are rated to carry current up to 60 A.

Norman Williams, K6BI
San Rafael, California

"Tion"

Dear Editor:

Which is often the tail end of a "conversation," and is sometimes baffling enough to inspire me to waste time guessing what might have preceded it, mind-boggling or otherwise. No need, I thought, for such mental exercise in the case of Dr. Belrose's four-column forest of words in the Winter issue of *Communications Quarterly*, for he thoughtfully provides his "return again to the basics" as the readers' launch-pad, by quoting the IEEE.

Unfortunately, excerpt "2" from the IEEE Dictionary proved baffling anyway to one who had based his grasp of radio tube thermionics post-1944 on Terman's *Radio Engineer's Handbook* (1943). What possible value could there be here in quoting the IEEE's resurrection of a definition that was falling out of favor before World War II?

I grant that Dow (1937) writes on page 267 of his electronics *opus* "*an output resistance* or *load resistance*"; but he then seems careful himself not to make use of the form "output resistance." I had always assumed that Terman (1943) wrote with lasting authority in his section headed "Output Impedance of Vacuum Tubes" on page 475, which clearly defines this variable as the impedance offered by a tube to an external applied voltage, a definition that is consonant with the IEEE excerpts "1" and "3," but not with excerpt "2."

If that desirable feature "return to basic" is to be interpreted as including in any three definitions one that is obsolete and contradictory, then, Ms. Editor, will you be advising contributors to interchange the "+" and "-" symbols on every third circuit schematic to conform with the convention on electrical polarities that held 100 or more years ago?

Desmond Thackeray, Ph.D.
Byfleet, Surrey, England

Lost and found

This letter, dated May 30, 1998, from W6ZVV must have fallen out of the mailbag. Now he and K1ZJH have picked up the discussion.

Dear Editor:

This is regarding the article by Dean W. Battishill, W5LAJ, in *Communications Quarterly*, Spring 1998, "The Care and Feeding of the 4CX1600B," on pages 91–97.

Figure 2 shows an incoming 240-VAC line having three wires: two hot wires ("red" and no color) with a "neutral" wire connected to earth ground (symbol) at the load.

It seems to me that when "...the editors [took] the liberty to redo the design using 220-volt components..." the "neutral" was no longer required as a common return and should have been replaced by a "ground" wire connected to the chassis of the load.

This circuit should be clarified, even in a conceptual piece; there are safety concerns about electrical shock and fire. A 240-VAC one-phase supply has a two-pole, three-wire grounding circuit: two hot wires (L1 and 2; black and red) with a grounding wire (G; green) connected to the load side.[1,2]

Please verify these references with their latest editions.

Details of the safety concerns have been discussed by Richard M. Lorenzen, WAØAKG.[3]

Thank you for your attention to my comments.

Hal Jones, W6ZVV
San Mateo, California

REFERENCES
1. *IEEE Recommended Practice for Grounding of Industrial and Commercial Power Systems*, ANSI/IEEE Std. 142-1982, "Green Book" (New York: IEEE, 1982), 83, Figure 24 (b).
2. *Guideline on Electrical Power for ADP Installations*, Federal Information Processing Standard Publication No. 94 (Washington: U.S. Dept. of Commerce/National Bureau of Standards, 1983), 18, Figure 7B/C.
3. *Ham Radio*, October 1988, "Comments," page 6.

W6ZVV responds to K1ZJH

Dear Peter:

This is regarding your clarification in *Communications Quarterly*, Spring 1999, page 4, to the article on W5LAJ's power supply design in the Spring 1998 issue. Last year I submitted comments to the Editor (enclosed) on that article, but never received an acknowledgment, nor have I found a published correction. I am gratified to see your attention to those safety concerns. Please consider the following additional remarks:

1. It was stated that "The Neutral should never be tied to the chassis ground under any circumstances."

The ground symbols used in the schematic of **Figure 1** are *earth* grounds, not *chassis* grounds (Ref: *ARRL Handbook*, 1994). It appears that the schematics in *Communications Quarterly* are not consistent in this regard. Perhaps some computer-generated schematics do not observe this convention. May I suggest that consideration be given to following contemporary electronics and communications standards, such as ANSI/IEEE "Standards for Graphic Symbols for Electrical and Electronics Diagrams."

2. You asked about the use of a four-conductor plug and socket:

The Bryant "Wiring Devices Catalog" BDB 1501, dated 11/81, shows the multitude of devices in the NEMA Configuration Chart. There are two major classes:
• Straight Blade Devices (15, 20, 30, 50, 60 ampere)
• Locking Devices (15, 20, 30 ampere).

Both classes have three-pole, four-wire grounding devices for 125/250 VAC, on line(s) 14; they are included with the three-phase devices.

I'm not a practicing electrician either; I suggest consulting a professional about a device for this application. See also "Electrical Safety," W6JIC, *QST*, April 1999, page 81, paragraph 2.

3. The power supply uses 240-VAC Variacs. I believe it has been standard practice to fuse the variable-output circuit, protecting the brush from overheating in the event of an overload.

Perhaps an electrical guru can confirm this.

Hal Jones, W6ZVV
San Mateo, California

Peter's Comments

Dear Hal:

Thank you for your concern regarding the W5LAJ power supply shown in the Spring 1999 issue. I am somewhat embarrassed about the "neutral" instead of "ground" being shown in the original Spring 1998 schematic, defeating most of what we had attempted to correct.

I concur with fusing the Variacs; any device that may be damaged by currents allowed by prior fusing should also be fused accordingly. The schematics are drafted using Ivex WinDraft software and the supplied component libraries. I'll check to see if upgraded libraries are available.

Peter Bertini, K1ZJH
Senior Technical Editor

On polyphase AM

Dear Editor:

I would like to thank you for publishing Warren Bruene's article "Polyphase AM" in the Spring 1999 issue. Readers may find it interesting that a version of the polyphase technique is used to provide the basic modulation components used for ILS and VOR navigation aids around the world.

The methods used to generate the modulation are similar to those described in the article, with some exceptions. Navaid systems generate independent DSBSC signals, which are called the SBO (sideband only) signal, and the carrier signal, which is called the CSB (carrier sideband).

Space modulation is the term used to describe an amplitude-modulated signal in which the modulation occurs in the far field, just as described in the article. However, in the navigation aids, because of the way in which the signals are applied to the antenna system, the modulation index of the received signal varies according to where the receiving antenna is in relationship to the transmitting antenna array.

The polyphase system diagrams in the article have many striking similarities to a modern VOR. A VOR or VHF omnirange provides pilots with a bearing to or from the VOR site. From the similarities in the diagrams, I would not be surprised to find that VOR's are a direct descendant of polyphase AM research.

A VOR uses a 30-Hz modulation frequency and generates a CSB and two SBO signals, with the SBO components locked in quadrature. The VOR system uses two pairs of antennas; a NE-SW pair and a NW-SE pair. One pair receives CSB and the sine SBO, while the other receives CSB and the cosine SBO. The resulting space modulation also generates a rotating Limacon pattern. If anyone has ever listened to a VOR signal, the 30-Hz warble heard in the modulation is a direct result of space modulation.

The addition of a 30-Hz reference signal modulated on a 9960-Hz subcarrier provides a phase reference. The VOR receiver measures the phase difference between the 30-Hz reference and the 30 Hz extracted from the space modulation, allowing the pilot to determine the azimuth to or from the VOR site.

An ILS, or instrument landing system, uses a localizer system to provide left or right of course information and a glide slope system to provide height above or below glide path information. The key modulating frequencies for both systems are a 90-Hz and 150-Hz signal, and the CSB signal is modulated with the same information present in the SBO.

When an aircraft is exactly lined up on course and on glide path, the modulation index of the 150-Hz component rises and the index of the 90-Hz component decreases. The ILS receiver measures the difference in depth of modulation (ddm) of the two signals and displays the difference on an indicator meter in the cockpit.

Not all technologies survive the competition to reach the market, but often pieces of what is learned along the way do. I do not know if, when, or how the polyphase process migrated to use in navigation aids, or if it was independently developed, but descendants of polyphase AM are definitely alive and well. Thanks to Mr. Bruene for an interesting article.

Dan Wanchic, WA8VZQ
St. Cloud, Minnesota

M.E. Gruchalla, P.E.
Member: Editorial Review Board
4816 Palo Duro, NE
Albuquerque, New Mexico 87110

IMAGE SCANNING

How much resolution is enough?

D igital photography and scanning are among the fastest-growing photographic technologies, and many amateur radio operators like to share memories of DX trips and pictures of their stations via the Internet. The cost of equipment for digital image processing has dropped drastically in the past few years, as the resolution has increased.

What resolution do we need? It's said that there's no need to scan an image at much greater resolution than the output device is capable of printing. However, if our purpose is to digitally archive our images rather than simply scan and print them, we may wish to retain all the information in an image—independent of whether all that information can be reproduced with one of today's output technologies. In this paper, I examine the scanning and printing process primarily from an engineering rather than photographic perspective, to determine—at least to a reasonable estimate—the maximum usable resolution in scanning and printing of amateur negatives and prints to capture effectively all the information in the image.

This is not intended to be a rigorous analysis of the digital-imaging processes. The primary purpose is to provide insight into some of the basic elements of digital-photography from a technical perspective, and to explore the important digital and photographic parameters.

The photographic art

In the photographic art, the needed optical resolution is a function of a number of parameters, including such subjective parameters as viewing distance and the nature of the subject. For example, a much higher resolution may be required for the presentation of an industrial photograph of a refinery than would be required in a soft-toned portrait. Photography as a true
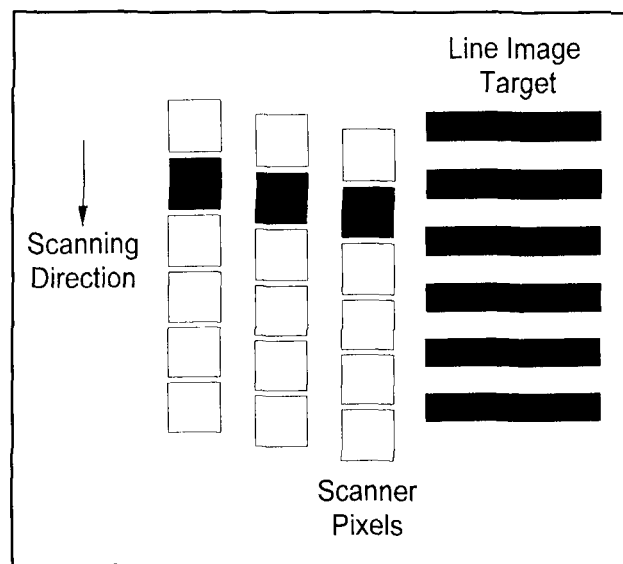
Figure 1. Scanner pixel size equal to target resolution (line pitch).

art form relies on the subjective artistic ability of the photographer to control such photographic parameters as contrast, color, and focus to produce a pleasing result for the viewer.

Photography as a science

Photography as a science is not subjective. Parameters of the various elements in photography as a science are specified analytically. Film resolution is precisely documented by such analytic tools as the Modulation Transfer Function. Film response to optical exposure is precisely provided by the sensitometric curve, also called the gamma curve, of the film.

In the science of photography, the intent is to eliminate virtually all subjective parameters by analytically measuring as precisely as one can as many of the photographic parameters as pos-
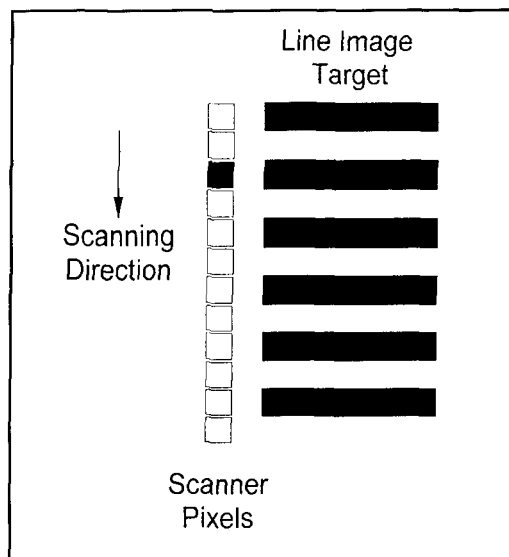
Figure 2. Scanner pixel size equal to target line width (one-half target resolution) pixel centered on black lines.

sible. The objective of this paper is to examine two fundamental analog materials of the photographic art—photographic film and photographic paper—and to determine from an objective technical perspective the equivalent limiting spatial digital resolution and optical digital resolution.

In general, the magnitudes of the values of these parameters from the technical perspective are expected to be much larger than those typically used in photography as an art. Objectively, the resolution is a measure of the limit of information storage of the analog medium. Theoretically, if an image were digitally recorded at spatial and optical resolutions equal to the limiting capability of film, the digital image would be virtually indistinguishable from an analog film recording (the statistics of the noise would be different, but the information content would be similar), regardless of such subjective parameters as subject content and viewing distance.

Digital photography and image scanning are precisely the same signal-processing process as digitizing an analog electrical signal. In processing typical electrical signals, care must be taken to assure that the sampling frequency is adequately above the highest frequencies in the signal to be digitized. If the sampling frequency is too low for the frequencies in the data signal, peculiar aliasing artifacts will result. Often the signal is prefiltered to assure that any signal content above some maximum frequency is sufficiently low to avoid unwanted aliasing. These same concerns apply in digital imaging.

We typically digitize an electrical signal in two ways, or two degrees of freedom or dimensions: We digitize the signal amplitude as a

function of time, and we digitize time by digitizing the signal only at discrete time intervals. The digitizing of an image, whether the image of a scene being photographed with a digital camera or a photograph scanned with a scanner, is precisely the same signal-processing process as digitizing a temporal signal. We are simply digitizing in the spatial rather than the time domain. Also, in the spatial domain, we are digitizing in five dimensions rather than the typical two for electrical signals. As with an electrical signal, the actual signal parameter of interest is the magnitude of the signal—the intensity of the image at various points. In order to identify where each intensity measurement is taken, both the X an Y coordinate positions must be recorded with each sample. It is convenient that both the X and Y position of each sample be digitized.

Generally, an electrical signal has only a single feature of interest: magnitude (there may also be interest in phase, but this is related to the timing in the sampling system with respect to some timing reference). When we digitize the "magnitude" of an element of an image, we don't want just a single magnitude, we want the magnitude of each individual color in the image element. Fortunately, we need only record the magnitude of three primary colors, red/green/blue or cyan/magenta/yellow, to record the full color of the image element. Therefore, in addition to digitizing the X and Y coordinate positions of an image element, we also digitize the magnitude of each of the primary colors in the image element. Therefore, we need five digital parameters to completely describe each individual element of an image.

Typically the magnitudes of these parameters are stored in a particular sequence, so the actual X and Y coordinate values associated with each element need not be recorded. For example, when an electrical signal is digitized, we do not actually record the time information. We simply record each digitized sample in order. Because we know the sampling frequency, we know the time spacing between samples and, in turn, the actual relative time that each sample was taken without actually having to record each time value. The same is true in spatial sampling. Because we know the sampling spacing in the X and Y directions, we need only record the magnitude information for each image element in some known sequence in order to keep track of where each piece of digitized magnitude information belongs in the image.

The position sampling or "spatial sampling" of an image is exactly the same as "temporal sampling" of an electrical signal. However, unlike electrical signals, we can't typically "filter" the incoming image signal. After all, how

does one filter the "bandwidth" of the image of a scene that is "input" to a digital-camera lens? If we aren't careful in selecting the "spatial sampling frequency" at which we sample our image, we could end up with the same aliasing artifacts in our digital image that can result in digitizing an electrical signal if the sampling frequency is too low for the frequency content of the signal.

To eliminate aliasing, we must assure that the sampling frequency of our digital camera or scanner is sufficiently higher than the maximum spatial frequency content of our image. The frequency of an electrical signal is specified as the number of magnitude cycles per unit time; i.e., cycles per second or Hertz. In the spatial domain, the spatial frequency is specified as cycles of intensity per unit length; i.e., cycles per millimeter, cycles per inch, etc. The image frequency in the X direction will generally be quite different from that in the Y direction. Also, the sampling frequency in the X direction may be different from the sampling frequency in the Y direction; consider a 300-dpi x 600-dpi scanner, for example. But, sampling in the spatial domain and the electrical domain are precisely the same, and all the same mathematics apply. Just the number of dimensions and the units are different.

Resolutions

There are two different resolutions that must be considered in digital imaging: spatial and intensity. Spatial resolution is a measure of how many discrete spatial positions the image is divided into per inch, millimeter, etc. in the *horizontal and vertical dimensions*. The spatial resolution is typically specified in dots per inch, or dpi. The higher the spatial resolution, the higher the number of positions that can be resolved.

The other type of resolution is optical-intensity. Because the intensity of the digital image is digitized, the original continuous intensity of the image is "quantized" into some number of discrete levels as in any digitizing process. The intensity resolution is typically specified in bits, referring to the precision of the digitizer used in the digitizing process. So, if the intensity of a photographic element is digitized to eight bits, a total of 256 intensity levels (255 levels plus zero level) are resolved. Unfortunately, as noted below, both the spatial resolution and the intensity resolution are confused a bit by the way manufacturers specify their equipment.

Scanning and printing

In general, the scanning and printing processes are considerably different due to the basic
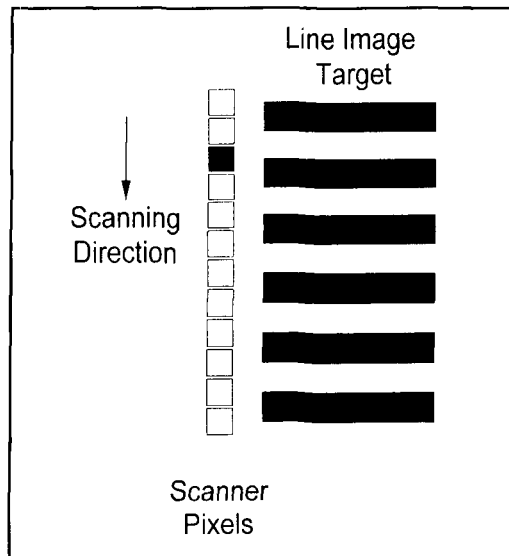
Figure 3. Scanner pixel size equal to target line width (one-half target resolution) pixel centered between black and white lines.

operation of the typical scanner and the typical "photo-quality" printer. The resolution of both scanners and printers are usually given in terms of dots per inch in horizontal and vertical directions, but this specification has a slightly different meaning for a scanner compared to a printer. Each dot a scanner digitizes is essentially a "scanner pixel." The pixel is the smallest picture element that can be resolved.

When the scanner digitizes a pixel, it digitizes the average intensity of each of the primary colors in that pixel. If the scanner is a "24-bit" type, it digitizes each color to eight bits of intensity resolution. The three eight-bit digitizing resolutions corresponding to the three colors are added to obtain the 24-bit specification. Specifying the unit as an "eight-bit" color scanner, where it is understood that each color level is digitized to eight bits, would provide the same information, but the 24-bit number is a much more impressive marketing figure. Similarly, a 30-bit unit digitizes to 10 bits and a 36-bit unit to 12 bits. But anything over 8 bits is not quite what it seems—more about that later.

So, the scanner "sees" a pixel as a dot, more or less, and digitizes its color intensities. A 24-bit scanner can see 256 different levels of each color intensity in a single pixel "dot." However, a printer like an ink-jet type can only print or not print a dot. It can't print part of a dot, or dots of different optical densities. Some printers do provide "image enhancements," and others such as dye sublimation printers can provide better continuous tone in each dot. But, because the ink-jet type printers are the most common, these are the only printers considered here.

To print an image that has a density somewhere between white (assuming a white paper

Line Image Target

Scanning Direction
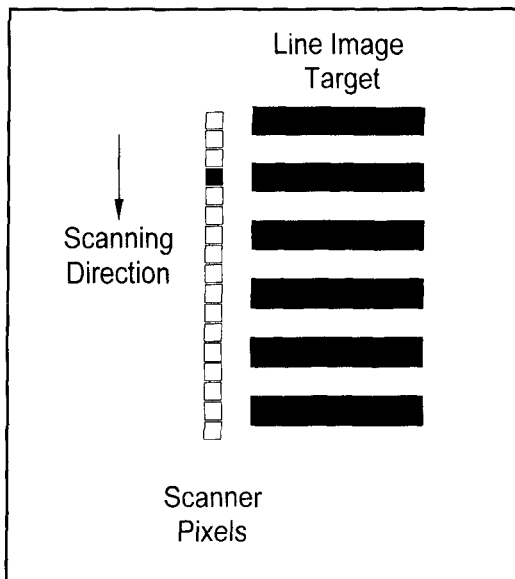
Scanner Pixels

Figure 4. Scanner pixel size equal to one-third target resolution pixel centered on black lines.

printing medium) and maximum density, the printer must print some dots and leave out some dots in a printed area to achieve a visual effect of a midrange intensity density. This is the basic principle used in half-tone printing. The result is that a "printer pixel" must have many dots in it to allow the printer to print pixel densities from total white to maximum density. I'll go into more printer detail later.

The Modulation Transfer Function

One of the critical characteristics of an optical device or material is its ability to resolve the

Line Image Target

Scanning Direction
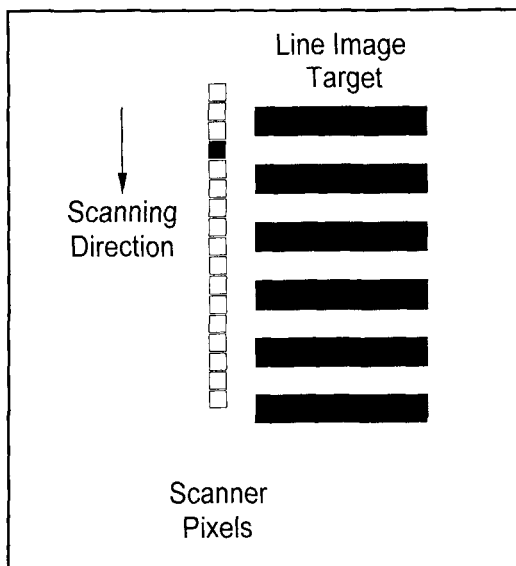
Scanner Pixels

Figure 5. Scanner pixel size equal to one-third target resolution pixel centered on white lines.

fine detail of a subject. As the detail of a subject becomes more fine, optical systems lose the ability to resolve the contrast of the subject. The Modulation Transfer Function, or MTF, is a parameter used to specify the spatial resolving ability of an optical element as a function of frequency. The MTF is a type of "photographic frequency-response" parameter.

Photographic resolution is typically specified in line pairs per millimeter. A line pair consists of a black line adjacent to a white line, where the black and white lines are the same width. When a number of such line pairs are placed together side by side, we see a series of parallel black lines separated by white spaces. The white spaces are actually the white lines of the line pairs. Because a black line plus a white line constitutes a line pair, the spacing or period or pitch of the line pairs is measured from one line of a line pair to the same line in the next pair. For example, the distance from the center of the black line in one pair to the center of the black line in an adjacent pair is the line-pair period. The unit of the line-pair period is millimeters per line pair (mm/lp). The reciprocal of this period is the resolution or frequency in lp/mm.

A test target is typically used to determine the photographic (as opposed to digital) resolution of photographic equipment and materials. Such test targets include a number of different standard patterns, including a number of line pairs of different periods. The test target is photographed and the processed image is examined to determine the line pairs with the smallest period that can be seen distinctly. This may be applied to both negatives and prints, and even to lenses where an aerial image is examined with a microscope.

"What can be seen distinctly" is a somewhat subjective measurement. What one person may be able to see, another may not. Or, if given an entire image, one may be able to "just see" more detail than if only given a very small part of the image containing fine detail. I will use a type of "half-power point" as my measure of the resolving ability of photographic materials.

The density of an image element on film is proportional to the intensity of the image and the time of the exposure. Because the optical intensity is power per unit area and the image element is an area, the image density is proportional to power per unit area times the image-element area times the exposure time. The product is energy in watt-seconds. When we photograph a subject, the exposure time of the entire frame is the same for all points in the frame. If the exposure time is constant for all points on the film, and only the image intensity at each point varies, the image density at each

point on the film is proportional to the "input" optical power at each point.

I'll use the half-power point of the developed image as the limit of the useful resolution of photographic film and paper. The MTF is a type of measure of the power bandwidth of photographic materials, and my half-power point is simply the "optical frequency" at which the MTF drops to a value of 0.5. In an electrical circuit, the half-power point is the frequency at which the response drops 3 dB from the mid-band value. However, the response is not zero beyond the half-power point; it simply drops with increasing frequency. So, there is still useful information beyond the half-power point. The half-power point or its equivalent is simply a convenient reference point for many physical phenomena. My choice of an MTF value of 0.5 as the limiting value is also simply a convenient reference point and one consistent with the electronic art. There's certainly resolution information at spatial frequencies above this point, but the density ratio becomes frequency dependent above that point—just as gain in an electrical system becomes frequency dependent above the upper −3 dB frequency.

The MTF is normalized to the contrast of the actual subject being photographed. Consider a line-pair test target containing line pairs of different frequencies used as the subject of a test phonograph. First, the contrast ratio of the subject is determined; for example, a particular line-pair target to be examined. This is called the Modulation M of the subject. Optical modulation is defined as the ratio of the difference of the maximum intensity minus the minimum intensity divided by the sum of the maximum and minimum intensities. The modulation of the subject pattern, the object of the photograph, is called M_O. The test object is photographed and the film processed. The image of the object is examined and the modulation of the image, M_I, is computed in the same manner.

$$\text{Object Modulation} = M_O = \frac{I_{O,MAX} - I_{O,MIN}}{I_{O,MAX} + I_{O,MIN}} \quad (1)$$

$$\text{Image Modulation} = M_I = \frac{I_{I,MAX} - I_{I,MIN}}{I_{I,MAX} + I_{I,MIN}} \quad (2)$$

The MTF is the ratio of M_I to M_O. Typically the object for MTF measurements is a test target designed with a constant modulation, perhaps 35 percent, and with a varying frequency. In other words, the target contains numerous line pairs of different periods all with the same contrast ratio.

$$MTF = \frac{M_I}{M_O} \quad (3)$$

One particularly useful test target is a sequence of line pairs where the period of the lines decreases across the target. In other words, the frequency increases across the target. By using such a target, the MTF as a function of image frequency may be very easily measured. This then provides the frequency response of the photographic element being tested.

A very good professional lens may have an aerial resolution of 300 lp/mm. A fine-grained film optimally processed may provide as high as 100 to 200 lp/mm resolution.[1] These are pretty much the upper end of what is available with standard techniques.

On a more practical level, a color negative produced with a camera of modest quality and competently processed by a bulk processor will provide a resolution of about 50 lp/mm or perhaps slightly higher.[2,3] In a color print, about 10 lp/mm or a bit higher is typical. Here, I consider only color photographs, as color is far more popular than black and white. Therefore, looking at the camera, lens, and film as a single system, I'll use a resolution on the order of 50 lp/mm in a color negative or slide as about the upper limit of what's achievable in the consumer market. This corresponds to 1,270 line pairs per inch (lp/in). For prints, I'll use 10 lp/mm, or about 250 lp/in, as the maximum typical resolution.

Scanning

First we'll look at scanning in a little more detail. What resolution is needed in a scanner? That depends on the object to be scanned. A photographic negative will require much higher resolution than a photographic print, and a magazine picture much less than a real photograph. Also, it is very likely that we may wish to "enlarge" a scanned image of a negative just as we would enlarge the image in a negative for a photographic print, so we'll need a higher resolution in the scanned negative image than in the final printed image. The best we can do is to capture all of the information available in whatever we are scanning. Let's look at the negative first. In general, this applies to slides as well.

Suppose we photograph a line-pair test target such that the image of the target on our negative is 1270 lp/in as noted above. Now, suppose that this target image is just visually resolvable in our processed negative; i.e., it's at the half-power point of the MTF of the film. This photographed image will now be used as our scanner test target. For convenience, I'll call the maximum-density lines of the target black lines "black" and the minimum-density lines of the target white lines "white." If we
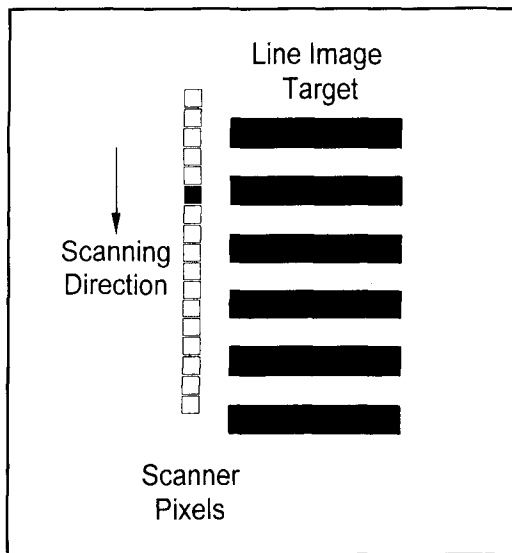
Figure 6. Scanner pixel size equal to one-third target resolution pixel positioned at the edge of black lines.

were to scan this target, what spatial scanning resolution would be required to just resolve this target image?

Test target 1

It might seem at first that 1270 dpi would be adequate; but, suppose we have a scanner with a spatial resolution of 1270 dpi. The pixel size is then 1/1270 pixels per inch or about 0.0008 inch. Then suppose we place the target in the scanner with the test-pattern lines running exactly at right angles to the mechanical scanning direction, so with each step of the stepper motor the scanner advances down the line pattern from one line to the next. We could orient the pattern at right angles to the detector array as well, but it's easier to visualize the stepper stepping along the line pattern. Finally, we carefully adjust the target so the center of each scanning pixel of the 1270-dpi scanner coincides exactly with the center of each black line in the 1270-1p/in test target of the negative.

Now, recall that a line pair consists of a black line and a white line side by side. For the 1270-1p/in target, the spacing between black lines is 0.0008 inch, so the actual black and white lines are each 0.0004 inch wide. However, the scanner pixel is 0.0008 inch; therefore the pixel spans precisely two entire lines, one black and one white. No matter how we adjust the target with respect to the scanner, as long as we do not rotate it, a pixel will always span exactly one black line and one white line (see **Figure 1**). The active pixel in each of the figures is shown filled black. The three columns of pixels in **Figure 1** represent three positions of the scanner with respect to the line target. Notice that

no matter how we slide the pixels with respect to the target lines, each pixel always spans exactly a black and a white line.

For example, for the first positioning (column of pixels on the left in **Figure 1**) the active pixel spans a white line, a black line, and another white line, so this pixel sees one whole black line and one whole white line. For the second positioning, the active pixel spans one whole black line and one whole white line. Finally, at the third position, the active pixel spans a black line, a white line, and a black line, so again the pixel sees a complete black line and a complete white line.

By pixel definition, the scanner cannot resolve any detail within the pixel. It can only resolve the average intensity of the entire pixel. So, if the scanner pixel spans both a black line, which is a minimum-intensity line, and a white line, which is a maximum-intensity line, the average intensity of the pixel is half maximum intensity, or 50 percent gray. And, if no matter how we adjust the target along the scanning direction a pixel always spans both a black and a white line, the entire image will always be scanned as 50 percent gray. We won't be able to see the individual lines of the target in the scanned image. Consequently, a 1270-dpi scanner resolution can't resolve a 1270-1p/in target. Instead of just resolvable black and white lines, a uniform gray image will be the result of the scan.

Test target 2

Now suppose a scanning resolution of 2540 dpi is used. In this case the pixel size is exactly the same as the width of the lines, or about 0.0004 inch. In general sampling theory, a periodic signal must be sampled more than twice per signal cycle to just be able to resolve the signal. If the scanner pixel size is exactly the width of a line, then exactly two samples per cycle of the line-pair target will be provided. Again, we position the target so a scanner pixel coincides exactly with the lines of the target. This is shown in **Figure 2**. Now, one pixel falls exactly on a black line and the adjacent pixels on either side fall on white lines. Therefore, as the scanner steps down the target, it will scan alternate black and white lines. It appears that a 2540-dpi scanning resolution is adequate to resolve a 1270-1p/in image, but this is not quite true.

Suppose we now move the target down along the scanning direction a distance of exactly 1/2 pixel, as shown in **Figure 3**. Now each scanner pixel will always span one half a black line and one half a white line. In **Figure 3**, the active pixel spans one-half a black line and one-half a white line. On the next step, the scanner will

move one entire pixel, so the pixel will now span one half a white line and one half a black line. In each case, the average pixel intensity is gray. Here, with the target adjusted so the scanner pixel falls halfway between black and white lines, the entire image will again be scanned as continuous gray. So, a scanning resolution of exactly twice the image resolution isn't quite good enough. This example (though perhaps somewhat simplistic) demonstrates why sampling a periodic signal more than twice per signal cycle is required to just resolve the signal.

Test target 3

Finally, suppose we have a scanning resolution that's a factor of three finer than the image resolution. This 3X scanning-resolution factor provides three pixels in the period of each line pair of the test target. For our 1270-1p/in test target, this is a 3810-dpi scanning resolution—a pixel size of about 0.00026 inch. Theoretically, a scanning resolution somewhat more than twice the image resolution could be used, but a factor of three is very easy to demonstrate here.

Again, we adjust the target so the scanner pixel falls exactly on the center of a black line as shown in **Figure 4**. In this case, the pixel is smaller than the target line so we could actually miss it a little and still be totally on the line. At this point, the pixel is digitized as totally black. When the scanner steps one step, the pixel will fall between the black and white lines with one third of the pixel on the black line and two thirds on the adjacent white line. The average pixel intensity will be digitized as about 33 percent gray.

When the scanner steps a second step, the pixel will fall two thirds on the white line and one third on the next black line, so again the average pixel intensity will be 33 percent gray. And, with a third step, the scanner pixel again falls totally on the next black line. So, again, we can resolve the individual black and white lines of the test target if we carefully adjust the position of the target with respect to the scanner. Of course, we should have expected this based on the previous example.

The scanned representation of our image is a series of black and gray lines. At the starting position the pixel is totally black, at the first step it is about 33 percent gray, at the second step it is also 33 percent gray, and at the third step is back to black again. Our scanned image appears as a series of black/light gray/black lines. The scanned image isn't quite the same as the original object, but what we are looking for is the ability to "just resolve" the individual black and white lines of the line pairs.

Now move the test target down one pixel width as before, or about 0.00013 inch, as
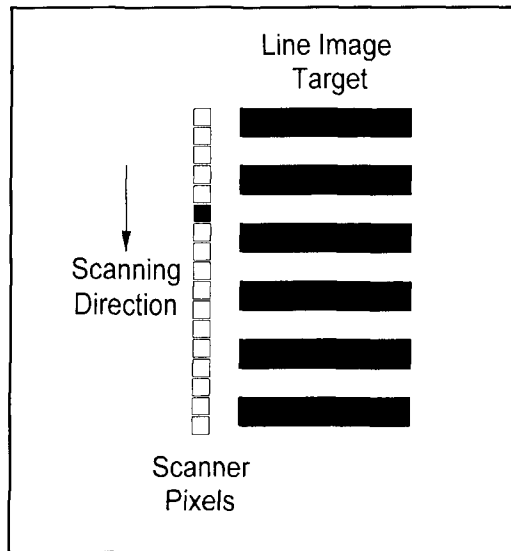
Figure 7. Scanner pixel size equal to one-third target resolution pixel positioned at the edge of white lines.

shown in **Figure 5**. Now the scanner pixel will alternately fall exactly on a white line and then part way between lines. Again scan three steps, but start on a white line. At the starting position the pixel is white, at the first step it's 66 percent gray, at the second it is also 66 percent gray, and the third step is white again. Now our scanned image appears as a series of white/dark gray/white lines. We can still resolve the image, but this image is different from the one obtained with the target adjusted for the scanner pixel initially falling exactly on a black line. Is this what we expected? Is it good enough?

This phenomena is an optical contrast reversal of the image caused by a type of aliasing in the signal processing of the digitized image (the image is discrete black and white lines) and the digital positioning of the scanner (the scanner can only position in discrete steps). This contrast reversal in the image of an object, such as a line-pair target, is an objective point of reference. In other words, it is an optical phenomenon that anyone can see with proper training, and there's little argument as to whether the contrast is reversed or not. Therefore, this point of contrast reversal can be used as a well-defined reference limit of image resolution expressed in 1p/mm, or 1p/in if you prefer.

Now, suppose we take the last example above with a scanning resolution of 3810 dpi, and position the target so the scanner pixel is totally on a black line but right at the edge of the line in the scanning direction, as shown in **Figure 6**. The pixel is black at this initial position. At the first step, the scanner pixel is then totally on the adjacent white line, so the pixel is totally white. With a second step, the pixel is exactly between lines so it's one-half black and

one-half white, and the average pixel density is 50 percent gray. With a third step, the pixel lies totally on the next black line, so the pixel is black. The scanned representation is a series of lines with a black/white/50-percent gray/black sequence. With this positioning of the target, we can still just resolve the target image.

If we move the target so the initial pixel is just on the edge of a white line, as shown in **Figure 7**, the pixel at this initial position will be white. At the first step it will be black, at the second step it will be 50 percent gray, and finally at a third step it will be totally white again. The scanned image will be a sequence of lines white/black/50 percent gray/white. As a result, there is still a contrast reversal because, in the first case, the sequence is black/white/gray/black and in the second it is white/black/gray/white. We could also view the first as gray/black/white/gray and the second as gray/white/black/gray. Such a contrast reversal is also called "pseudo resolution," as it produces apparent additional lines between the real lines.[4,5] In this example, we can easily see that the pseudo resolution has provided a third virtual line in the scanned image. There's no gray line in the actual target between any of the black and white lines.

Test results

In the three test cases examined here, it appears that a 3:1 scanning-to-image resolution is adequate to provide a scanned image that can just resolve the source image, but will provide a contrast reversal. At this scanning resolution, the ability to just resolve the image is not dependent on the positioning of the target in the scanner. Of course, the scanned image is different for different positions, but we can always just resolve the target image. In general sampling theory, the Nyquist Frequency is a factor of two above the frequency of the signal being sampled, whether it is an electrical signal or a photographic line-pair test target, and the sampling frequency must be greater than the Nyquist Frequency to preserve the signal information and to prevent aliasing.[6]

Perhaps a scanning resolution a factor of three greater than our negative resolution is what we need at a minimum. But, even at that resolution our line target is not really faithfully reproduced because, although we can just visually resolve the target on the film image, we will see a contrast reversal and pseudo resolution as a function of how we position the target in the scanner. This may result in Moire patterns in images that contain very fine detail consisting of parallel lines. Also, I have used the half-power point of the MTF as a point of reference. There is certainly additional image information beyond the half-power point. A scanning resolution that's nominally a factor of three finer than the maximum image spatial frequency will only provide about one bit of information about the image at the maximum image frequency—whether anything is there or not.

Generally, a photographic resolution of about a factor of five higher than the maximum spatial frequencies found in the subject to be photographed is considered the typical resolution necessary to capture virtually all the detail of that subject. This is also pretty much true for scanned images. It's equivalent to providing an electrical bandpass in a system that's substantially higher than some electrical signal frequency of interest in order to prevent the bandpass of the system from altering the signal. But, as I'll show later, the higher the scanning resolution, the greater the memory we'll need to store the image—the memory required goes up by the square of the resolution. The storage requirement increases by a factor of four if we simply double the scanning resolution. There is a serious compromise between how precisely we digitize an image and where we store it.

Examined on a microscopic scale, a photograph is composed of a sort of dot structure called "grain." Grain refers roughly to the individual silver-halide crystals that form the photographic image (this is not quite accurate for color materials, but the effect is the same). The grain size and spacing is a function of film speed as well as other parameters, such as storage and processing of the film. A medium-speed film may have a grain size and spacing on the order of a micron or so, or about 500 to 1000 grain elements per mm, about 25,000 grain elements per inch.[7] So, it's virtually impossible to truly capture with a scanner the actual granular information of the film. But such fine detail isn't actually necessary because the film itself is not capable of providing spatial image resolution equivalent to the grain spacing. As noted above, the MTF of a typical color film limits the spatial resolution to about 50 lp/mm, or about 1270 lp/in.

Fortunately, nature is kind to us in typical photographs. Rarely do we wish to photograph resolution charts, such as line charts or pie charts, to proudly display in our home gallery. I focus in this paper primarily on how much scanning resolution is needed to faithfully archive the image information of a typical amateur photograph or negative. In general, these photographs will be of people, pets, flowers, landscapes, and other generally random images—images that tend to have random intensity characteristics as opposed to very structured line characteristics, such as a line-pair test target.

These amateur photographs aren't necessarily of lower quality than professional photographs. They simply won't typically have the same subject content. Neither will these images typically have the equivalent of line-pair resolution test targets. For these typical types of images, a scanning resolution of about a factor of two greater than the image resolution is usually adequate. This isn't necessarily true of "professional" photographs where industrial products are often the subject; for example, a microphotograph of the chip of a Pentium II processor which will have very highly structured image characteristics very much like a resolution test target.

To reliably capture the full image information in a professional industrial photograph in a scanned image, a higher resolution of about a factor of five higher than the resolution of the photographic medium is required. The equipment involved will be much more costly than that required to capture the "useful" information in a typical amateur photograph. Of course, if you feel your photographs are of professional quality and of the types of subjects that demand the highest-possible scanning resolution, you may want to invest in professional-quality equipment, along with some means to store and print the resulting images. I am limiting my analyses to typical amateur photographs and negatives of random-density images.

Two film scanners available on the consumer home market are the Hewlett Packard PhotoSmart scanner and the Nikon CoolScan scanner. In negative/transparency mode, the scanning resolution of the HP unit is 2400 dpi x 2400 dpi and that of the Nikon unit is 2700 dpi x 2700 dpi.[8,9] This is pretty close to the estimated 2540-dpi resolution computed above and is required as a minimum to just resolve a typical amateur negative. This is a good compromise based on the available components, such as CCD imaging devices, positioning technology and the final digital-file storage requirements, and price.

Basic requirements

Recapping so far, a typical negative can provide a resolution of about 50 lp/mm or about 1270 lp/in. A scanning resolution about a factor of three greater than the film resolution, or about 3810 lp/in, is needed to reach the point of contrast reversal in a 1270-lp/in test target. And, if the full detail of the image is to be preserved, a scanning resolution of about a factor of five higher than the film resolution is needed, or about 6350 lp/in. But, because typical subjects of amateur photographs tend not to have highly structured detail expected in pro-

fessional industrial photographs, a scanning resolution only about a factor of two higher than the film resolution is usually adequate to capture virtually all the image information. The 2400 dpi provided by scanners like the Hewlett Packard PhotoSmart scanner is just at this minimum requirement. Scanners like the Nikon CoolScan provide 2700 dpi, which is just a bit above the minimum requirements. We'll look at this in a little more detail below in the review of storage requirements.

If a resolution of about 2500 dpi to 2700 dpi as provided in film scanners is just adequate to scan a photographic negative if all of the spatial detail is to be captured, what resolution is needed to scan a print? That parameter can be determined in the same manner as for the negative.

The resolution limit for a reasonable-quality, bulk processed print is on the order of 10 lp/mm, or about 250 lp/in. Using the same reasoning as above, we must scan such a print at a resolution of at least 500 dpi (2X) to capture all the useful spatial data in the print image—750 dpi (3X) if we are to reach the contrast-reversal resolution, and perhaps 1250 dpi (5X) to retain virtually all spatial information. A typical magazine photo is printed at about 133 dpi, so a scanning resolution slightly greater than about 266 dpi (or about 300 dpi) is needed to capture a printed magazine image.

Low-end scanners provide an "optical" scanning resolution (as opposed to interpolated) of approximately 300 dpi x 300 dpi. This is quite adequate to capture all of the spatial information in a typical printed image in a magazine or newspaper, but it's not quite adequate for true photographic prints. This does not mean that these 300-dpi scanners are useless or in any way substandard. We must decide what we need, or perhaps more appropriately what we want.

For the most part, magazine pictures look pretty good. If we are scanning prints for articles, letters, or other general interest applications, 300 dpi is quite adequate. But if we are scanning photographic prints to make electronic archives of them for digital preservation, or if we are scanning photographs that will be cropped and enlarged, we probably want to capture as faithfully as practical all the information available in the original print. To do so, we'll need at least a 500-dpi scanner, preferably 750 dpi. And, if we want to preserve essentially all the spatial information, we may need 1250 dpi. The medium-range scanners are 300 dpi x 600 dpi units, which is a reasonable compromise but still a little lacking in one dimension. The 600 x 600 dpi units are getting into the medium to high-end market, but the prices are coming down all the time. There are several 1200 dpi x 1200 dpi units available for about $300 to $500 for the basic scanner.

The unequal scanning-resolution specification for some scanners, such as 600 dpi x 1200 dpi, may seem a bit peculiar, but it is a result of how the scanners operate. In one dimension, typically across the narrower scanning dimension, the resolution is determined by the optical sensor, generally a CCD line array. This is simply a number of photodetectors side by side forming a single line of optical detectors.

The spatial resolution at the detector is determined by the spacing between the individual optical elements. The total width of the scanning area is imaged onto this line array. The scanning resolution is determined by the total number of detectors in the array divided by the width of the scanning area. For example, if there are 2500 elements in the detector array and the scanning width is 8.5 inches, the scanning resolution is approximately 300 dpi. An array slightly larger than 5000 elements would be required for 600 dpi. Because the number of elements in the optical-sensor array of the scanner directly relates to the cost of the optical array, the higher the resolution in the direction of the optical array, the higher the price of the scanner.

To scan the length of the image we move the scanning carriage along its length with a standard stepper motor assembly similar to that used in printers. The length of each step the carriage is moved is a function of the stepper-motor resolution (steps per revolution) and the effective gearing ratio used to couple the motor to the carriage assembly. This motion could also be provided by a linear motor, but it doesn't appear that this technology is as yet used in scanners.

Theoretically, the steps could be made as fine as desired, provided the stability of the mechanical system were adequate. For example, mechanical stepping is also used in computer disk drives to step between tracks (cylinders). Densities of almost 15,000 tracks per inch are used in modern hard drives—the Maxtor Diamond Max Plus Family of drives incorporates track densities as high as 14,522 tracks per inch for example.[10] So, generally, positioning technology is adequate to provide 1200 steps per inch in a scanner.

Initially, the scanner mechanical scanning resolution, that is the number of steps per inch, was made equal to the resolution determined by the optical detector array; a 300 dpi x 300 dpi scanner, for example. Then, for some reason, like marketing or improvements in the mechanical positioning technology of the scanning mechanism, twice the resolution in the mechanical scanning dimension became popular, such as the typical 300 dpi x 600 dpi units common now. This could have been a "no-cost" improvement provided by improved stepper-

motor technology, tighter process controls, or other cost-insensitive manufacturing or materials changes. If the doubling of resolution in the mechanical scanning direction could be provided at little increase in cost, it could be offered as an "improvement feature" at little additional price to make the product more attractive over competitive products.

If the pixel size in both the X and Y scanning directions are the same, however—that is the photodetector pixel is square—it's not clear that scanning in one direction at one half the pixel size is particularly useful. It's a type of "interpolation," that creates a synthetic pixel between each real pixel. This provides a type of spatial filtering of the scanned image. It is not clear that having twice as much resolution in one dimension provides much improvement in image quality. It's likely as much a marketing tool as a usable feature. And not all manufacturers are jumping to provide this feature; for instance, the Hewlett Packard 4C scanner is simply a 600 dpi x 600 dpi unit.

Interpolation

In most scanner specifications, we'll see both an "optical" resolution and an interpolated or "maximum" resolution. Sometimes the optical resolution is omitted because it's the smaller (but most important) number. The optical resolution is the actual physical resolution the scanner is capable of providing. The interpolated resolution is a figure that results from the scanning system (software and/or hardware) "creating" additional data points between the actual scanned real data points. This is not real data. It is simply data computed by the scanning system based on a mathematical algorithm.

For example, consider a scanned image that results in alternating black and white pixels and in turn alternating black and white lines. If the scanning system is allowed to "interpolate," it computes several additional pixels in between the true pixels. In other words, it makes the true scanned pixels smaller and adds a few more in between that aren't true pixels. With these added pixels, a smoother transition in density is provided, whether it was in the original image or not. Consider that the system provides a 4X interpolation in both directions, say 300 dpi optical to 1200 dpi interpolated. For each real pixel, it will compute three new ones along both scanning axes so there are 16 times as many pixels in the interpolated image—the original pixel plus the three computed ones in each direction.

Say that a simple linear interpolation between adjacent real pixels is used. For a black and white line target starting at a real

black pixel, the next pixel computed will be about 66 percent gray; the next, also computed, about 50 percent gray; the next, computed as well, about 33 percent gray; and finally the fourth pixel will be a real white pixel. So instead of a series of black and white lines corresponding to the original image scanned, the interpolated result is a series of more lines with a type of gray-scale transition between each real black line and each real white line. Clearly this does not add any information to the original image. It just makes the scanned image "look better," or not.

Basically what interpolation does is reduce the apparent pixel size, and this reduces the "pixelization" of the scanned image. For example, if we scan an image at 300 dpi and enlarge it enough, we will begin to see individual pixels. If the scanner interpolated by a factor of four, the pixel size would be one fourth as large; but the added pixels are not true image information. These pixels are simply estimates computed by a set of mathematical rules in the scanner system (software and hardware). As noted above, the grain size of typical film is much finer than the resolution that the film is actually capable of providing. So, at a microscopic level, because there are many more grain elements in the film per inch than resolvable lines, the grain itself provides a type of interpolation in the actual film.

Interpolation apparently is most useful where a small portion of an image is scanned and then enlarged. Here the finer pixelization provided by the interpolation reduces the jaggedness of the image, but the image will still be fuzzy because no actual image information is added by interpolation. It is not clear that interpolation is a useful feature in all applications. But we will get it "for free" whether we want it or not, although we do not have to use it. You will have to determine whether your images "look better" with or with out it.

Interpolation of an image does not necessarily increase the storage memory requirements of the digitized image even though the number of pixels may be increased by as much as a factor of 16 (4X interpolation in both axes) or even more. The interpolated information is computed from the real image data. Therefore, it may be computed at any time using image-processing software if we have the original digital image data available. Specifically, it does not have to be computed at the time of scanning. It may be just as accurately computed at the time of printing. Therefore, the basic scanned image may be stored without any interpolation, and at the time of printing the interpolation algorithms may be applied to provide an interpolated result in the printed image.

Not only does storing only the basic digitized image reduce the storage memory requirements, but it also allows us to apply whatever post processing, such as interpolation, we may desire at the actual time of printing to achieve a specific visual effect. For example, if we are printing the full-frame image in a large format to be viewed at a distance, we may not need any interpolation, but if we are cropping out a small part of the original image to be enlarged we may need a very high degree of interpolation to minimize pixelization in the printed image.

The basic scanned image is in effect the "raw data" of the original image. By storing this raw data, we avoid any additional corruption of the image (other than the fact that we digitized it spatially and digitized the intensity of its three primary colors) by such additional post processing as interpolation. If we always have the raw data available, we can then always apply new and wonderful post-processing techniques to that data to produce interesting printed subjects. However, once we apply some type of post processing to the raw data, generally the original raw data cannot be recovered from the post-processed result.

So, for archival purposes, it is most prudent to store the basic image from the scanner as faithfully as possible without any post processing, and do all post-processing image manipulation at the time of printing. If we wish to be able to reproduce a specific printed image at a later time, we need only record what we did to the raw data to obtain that print. This is precisely what a photographer does. When a negative is printed, the photographer does not in any way modify the actual negative. The negative is the photographer's raw data. The photographer will typically crop the enlarger image, not the negative, to the specific subject of interest and then "post process" on the image, for example by "burning in" dense areas of the image and "dodging" thin areas, to achieve the final desired visual results in the print. To be able to reproduce the print, the photographer simply makes written notes recording how the image was manipulated to achieve the final print, but never actually modifies the negative, the true "raw data."

Digitizing the image intensity

Now that we have a good idea of the spatial precision needed in the digitizing process, we can move on to the digitizing of the intensity information of the image. An individual can visually resolve about a 4 percent difference in the density of side-by-side samples. That is about one part in twenty-five. The dynamic

range of film is measured in terms of Optical Density, or OD. The OD ratio is 10log of the ratio of the darkest image the film can produce to the lightest image that the film can provide.[7] The lightest the image can be in a film image is the film base, and the darkest is the density of the emulsion at maximum exposure.

A typical film such as Kodak Royal Gold 100 provides an optical-density range of about OD 2 to 2.5 for each color.[2] An OD 2 is a density range of 100:1 and OD 2.5 is about 316:1. At either end of the exposure curve (sensitometric or gamma curve), the response of the film becomes nonlinear, so the actual usable range of the film is slightly less than the full OD range.[7] An OD of about 2.4 is a convenient usable density since it corresponds to a density range of almost exactly 256:1, or eight bits. Even though the eye can only resolve about a 4-percent intensity variation, typical film is capable of resolving about one part in 256. Therefore, digitizing each color intensity of a pixel to eight bits captures effectively all of the useful intensity information in the film image. As noted above, a scanner that digitizes each color to eight bits is designated a 24-bit scanner (3 colors x 8 bits per color = 3D 24 = bits).

Even though an OD of 2.4 may be the maximum range typically usable in a typical film image, there may be some additional information in the nonlinear areas of an image. For example, in over-exposed areas of the negative or in very thin underexposed areas, there may be some useful detail. As noted below, we would not typically be able to see these details when the negative is normally printed. Nevertheless, the information may be available in the negative. So, a higher digitizing resolution of the intensity would be useful. But if the eye can only see about one part in 25, why would we want to digitize a negative even to 256:1 much less even higher?

The reason is to capture all the information that is in the image, not just that which we may see in a print. If we archive all the information, we can use the additional information to optimize the output image. For example, when printing a negative by hand, one can burn in dense areas and dodge thin areas to bring out the details in those regions of the negative in the print. So if we capture all that detail in a digitized image, we will have the same optimizing opportunities in the digital image when we "print" the digital image with a digital printer.

Eight bits of intensity information is very near the limit of what the film can actually provide. This is a very convenient resolution for storage with eight-bit digital words. However, it is possible to digitize the image intensity to 10 bits or even 12 bits. There is a reasonable compromise among available intensity-digitiz-

ing resolution, useful intensity range of the negative, and limitations in digital storage. Some scanner manufacturers provide 10 and 12-bit scanning (specified as 30 bit and 36 bit, of course), but the scanned image is post processed after scanning, and only 8 bits per color are actually stored.

By applying a mathematical function to the digitized image (this has also been referred to as a gamma function), information in the dense areas may be lightened and information in the thin areas may be darkened. The gamma of a film, γ, is the slope of the plot of optical density as a function of the log of the exposure. This mathematical function may be used to effectively modify the gamma characteristics of the scanned image. It is typically directly manipulated by the user to "optimize" the scanned image. The additional digitized bits of information are used as additional data to make modifications to the scanned image. Effectively, the 10 or 12-bit intensity information of each color provided by the scanner is compressed by the mathematical function into an 8-bit intensity word for each color for final storage.

Printing

Printing of a digitized image is quite different from the digitizing process. Because the ink-jet-type printers are by far the most common and least expensive, I will consider only those in this article. However, you should be aware that there are quite a number of other printing technologies. Since typical ink-jet printers can print only fixed-sized dots of fixed density, continuous-tone information must be printed as half-tone fields of dots with different numbers of dots per unit area (different dot densities). Therefore, one could define a print pixel as being made up of some number of possible dot positions. If a dot is printed in each possible position in a print pixel, that pixel will be of maximum density. If no dots are printed in the pixel, the pixel is obviously of minimum density.

We must determine what range of density is needed in typical printed images. For the purposes of this paper, we will assume that the three individual dots of the three primary colors are printed at the same point. Each color of a print pixel is individually half toned by the dot density of that color in the printed pixel. Be sure to keep in mind that a scanner pixel and a printer pixel are different; the scanner pixel is the smallest dot that the scanner can resolve, but the printer pixel is typically made up of many printer dots.

One convenient place to start is to examine reproducing all eight bits of intensity information recorded for each color. This would

require 256 dots per print pixel, which can be provided in an 16 x 16 dot array. Earlier, I stated that a typical photographic print resolution is about 250 lp/in. If we have a printer that can print 250 dpi, it can exactly place 250 dots per inch in each printing direction. If the width and length of the dots are somewhat smaller than the dot and line spacing, a 250-dpi printer can just barely print 250 lp/in—each dot is part of a black line and the space between dots is a white line. Generally, the dots overlap somewhat so the printer is capable of printing solid colors without a discernible dot pattern. If adjacent dots just overlap, a slightly higher printer resolution will be required to reproduce the 250-lp/in line image. For a 250 lp/in line image, about a 300-dpi printer should be just capable of reproducing a 250 lp/in image.

It may seem that we do not need as high a resolution in a printer to print an image as we need in a scanner to capture an image. How-ever, the printer cannot print continuous-tone dots. So, although we may only need a printer that can print 300 lp/in, each printer pixel must be made up of many dots. For a 16 x 16 array, each pixel along a printed line must be broken into 16 dots, and there must be 16 lines per print pixel. Therefore, to produce a 300-lp/in image resolution with a 16 x 16 dot array for each print pixel required to reproduce an 8-bit gray scale, a printer with a 4800 dpi x 4800 dpi resolution is required. This is well beyond the present state of the art in typical ink-jet printing devices.

In the case of scanning, the goal was to record for archival purposes all the image information available in the negative or print being scanned. This required a scanning resolution consistent with the resolution of the negative or print. However, when we print an image, our primary purpose, if not only purpose, is to view the image. Therefore, we need only print at a minimum the information the eye can resolve. As noted earlier, the eye can resolve a density difference of about 4 percent, or about one part in 25. Therefore, if a 5 x 5 dot array is used for each print pixel, a 26:1 half-tone gray scale can be reproduced (one dot to 25 dots, plus no dots). This would require each pixel along a printed line to be broken into five dots, and five lines would be required for each print pixel. So, to print 300 lp/in with a 4-percent gray-scale resolution, a printer resolution of about 1500 dpi x 1500 dpi is needed.

This is just at the state of the art in ink-jet-type printers. Several printers are available that provide in excess of 1400 dpi x 1400 dpi color resolution. These printers should be capable of printing images that are virtually visually equal to a photograph in printed resolution. However, this only is true if these images are viewed as normal photographs are viewed. If viewed in close detail with a loupe or micro-

scope, the individual dots forming the half-tone scale will likely be easily seen. A photograph is a continuous tone image, almost. So, on a microscopic scale, a 1500-dpi printed image will be very much more course than a photographic image. Nevertheless, about 300 printer pixels per inch at 5 dots x 5 dots per printer pixel will reasonably reproduce the full visual range of sharpness, color, and contrast available in a typical photographic print under normal viewing conditions.

As noted, there are a number of other printing technologies available. Some of these, such as the dye sublimation printers, are capable of providing very high resolution, but typically at a much higher cost than ink-jet printers, both in the printer itself as well as in its consumables. Also, some of the newer ink-jet printers are using more than four inks (typical inks are cyan, magenta, yellow, and black) to provide some tone variation within each printed dot. This reduces the number of dots needed in a half-tone pixel to provide the desired density resolution (~4 percent) in the printed image.

Before you purchase a photo-quality printer, you need to decide what quality you need (want), and investigate the various printer technologies that will work the best for you. Be sure to check the price of the consumables, such as ink, toner, and dye-transfer film, before you settle on a printer. The cost of these consumables can quickly and easily exceed the price of the entire printer.

Storage memory requirements

One of the more serious limiting parameters of digital image processing is the memory requirement for storing the images. There are a number of compression algorithms that can very impressively compress digital image files, but generally with some loss of image information. For our purposes here, we will assume that no compression is applied since it is our purpose to exactly reproduce the full digital image which we capture.

You will remember that a 24-bit scanner provides three 8-bit bytes of pixel intensity information for each pixel. Even the 30 and 36-bit scanners typically generate three 8-bit bytes for each pixel through a software algorithm. So, for this analysis, we will use three 8-bit bytes to represent the full color intensity information of each pixel. The memory required for storing various image formats is computed below.

Consider film digitizing first. About the most common film format is the 35-mm format, and about the largest is the 4 x 5-inch format used primarily by professionals. There are of course a number of format sizes in between, but the two considered here will provide an upper and

lower limit of the memory requirements. A 35-mm negative is approximately 24 mm x 35 mm and the 4 x 5 negative about 102 mm x 127 mm. From above, the MTF of a typical color film is about 50 percent at a spatial resolution of about 50 lp/mm. If the negative is a professional image, we need to digitize at about a factor of five higher resolution than the actual film spatial resolution. As a result, for a 24 mm x 35 mm negative with a resolution of 50 lp/mm, we must scan at about 6000 x 8750 pixels. The total number of pixels is then 52.5 M pixels. And because each pixel contains 3 bytes of intensity information, a total of about 158 MB of storage is needed for a single professional 35-mm image. For a 4 x 5 negative, about 2.4 GB of storage is required.

Although these memory requirements may seem totally unmanageable, they are well within the current state of the art in storage media. For example, a single CD-ROM would store about four of these 35-mm images, and a DVD could store about seven 4 x 5 images. So, even though these are rather large files, it is well within the capability of presently available equipment to easily store them. Processing them however, may not be too convenient given the present state of the art in equipment and software.

From the contrast-reversal arguments explored earlier, we need a scanning resolution that is only about a factor of three higher than the image spatial resolution to reach this contrast-reversal reference point. Then about 57 MB is required to store a 35 mm image and about 871 MB for a 4 x 5 image. A CD-ROM can then hold about eleven 35-mm images and a DVD about nineteen 4x5 images.

Finally, if the images are typical amateur photographs, a scanning resolution that is about a factor of two greater than the film spatial resolution is more or less adequate. About 25 MB is required for a 35-mm negative and about 387 MB for a 4 x 5-inch negative. This is roughly the scanning resolution provided by presently available negative scanners like the HP and Nikon units referenced earlier. At this scanning resolution, a CD-ROM can hold about twenty-five 35-mm images and can actually hold one 4 x 5 image. A DVD could hold about forty-four 4 x 5 images and about 675 35-mm images.

If our purpose is to digitally archive our images to accurately preserve them and prevent degradation of the image information over time, the number of images that can be stored on a single medium is not too important. For example, even the limitation of storing only four 35-mm images on a CD-ROM at the maximum scanning resolution of a factor of five greater than the film MTF resolution is acceptable considering the current price of recordable CD-ROM media. At the 2X scanning resolution

Table 1. Memory requirements for various photographic resolutions (without compression).

Image scanned	Memory required
35 mm negative	
(50% MTF @ ~50 lp/mm)	
5X (6350 dpi)	158 MB
3X (3810 dpi)	57 MB
2X (2540 dpi)	25 MB
4 x 5 negative	
(50% MTF @ ~50 lp/mm)	
5X (6350 dpi)	2.4 GB
3X (3810 dpi)	871 MB
2X (2540 dpi)	387 MB
8 x 10 print	
(50% MTF @ ~10 lp/mm)	
5X (1270 dpi)	387 MB
3X (762 dpi)	139 MB
2X (508 dpi)	62 MB
11 x 14 print	
(50% MTF @ ~10 lp/mm)	
5X (1270 dpi)	745 MB
3X (762 dpi)	266 MB
2X (508 dpi)	119 MB

presently available from several film scanners, the capability of storing about 25 full 35-mm images on a CD-ROM is quite acceptable.

Now let's look at memory requirements for print digitizing. Although the typical resolution available in a print is much lower than film, about 10 lp/mm for prints as compared to 50 lp/mm for film, prints are typically very much larger. Most larger scanners provide a scanning area of 8.5 x 14 inch and the smaller units 8.5 x 11 inches. The closest standard photographic print size is 8 x 10 inches.

Consider an 8 x 10 print (203 mm x 254 mm) having a 10 lp/mm resolution. At a 5X scanning resolution, the scanner resolution must be 1270 dpi x 1270 dpi. Then, for a nominal photographic resolution of 10 lp/mm, a storage memory of 387 MB is required at 5X scanning resolution, 139 MB at 3X and 62 MB at 2X. Now consider an 8.5 x 14 scanner (216 mm x 356 mm) and a full 8.5 x 14 photograph. The memory requirement at 5X is 577 MB, at 3X about 208 MB and at 2X about 92 MB. Finally, for an 8.5 x 11 scanner (216 mm x 279 mm) and a full 8.5 x 11 print providing a 10-lp/mm spatial resolution, 452 MB of storage memory is required at a 5X scanning resolution, 163 MB at 3X, and 72 MB at 2X. An 11 x 14 print (which may be scanned in a 11 x 17 "tabloid" scanner) will require 745 MB at the maximum scanning resolution.

Tables 1 and **2** present the memory requirements for several photographic formats and scanners. No interpolation or compression is considered in these tables. If interpolation is used, the memory requirements will increase substantially, and with compression memory requirements may be reduced, but perhaps at the expense of some loss of image information.

Digital cameras

Finally, a brief comment or two on digital cameras is in order. A high-performance consumer digital camera presently provides an image resolution of about 1200 pixels x 1000 pixels (mega-pixel cameras). Note that this is not pixels per millimeter or per inch but rather total pixels in each direction. Recall that the negative scanners referenced above provide a resolution of at least 2400 dpi, or about 3300 pixels x 2300 pixels for a 35-mm negative, which is just below the limit of the scanning resolution required to capture most of the image information of the negative. So the high-end consumer digital camera provides an image area that is about one sixth the image area of a scanned 35-mm image (1200/3300 x 1000/2300).

In other words, the digital camera image is effectively a smaller image format than 35 mm. It is roughly equivalent to a 16-mm format, perhaps slightly larger. Therefore, the printed image from this digital camera must be printed about one-half to one-third as large (one third the length and width) as a print from a digitized 35-mm negative to provide the same spatial pixel density in the print.

So, just like in "analog" photography when using a "small-format" negative, when using a digital camera we must be much more careful to capture only the subject in which we are interested and fully fill the frame with that subject. And because the image is smaller, even if we fully fill the frame with the desired image, we cannot print it as large as a larger-format negative.

Digital cameras are still quite a way from being capable of reproducing the performance of typical 35-mm film cameras. Above we noted that a typical negative film can resolve about 50\~lp/mm and that we need a spatial digitizing pixel resolution of about a factor of three higher than the image resolution. So, in a 24 mm x 35 mm negative, there are 1200 x 1750 resolvable lines. Then, to adequately capture the same spatial information as the film is capable of capturing, the digital-camera resolution must be about 3600 x 5250 pixels (a digitizing resolution that is a factor of three higher than the film resolution). So, the digital cameras are about a factor of four away from 35-mm film cameras.

Basically two breakthroughs are needed to make the digital cameras equivalent to film cameras. One is higher-resolution CCD arrays and the other is high-density, portable digital storage media. As noted in **Table 1**, a memory capacity of about 60 MB is needed to store the digital information from each digitized 35-mm image if digitized at about 3X to capture all the information that 35-mm film is capable of providing. The Super Disk technology presently available provides 100 MB storage on a 3.5-inch floppy-format disk. There are also ZIP drives, optical floppy drives (floptical drives), and perhaps even other technologies that could be used as well to provide this capacity. With the continual advances in processing of solid-state components such as CCD arrays, it is very likely that a nominal 4000 x 5000 pixel device will be commercially feasible in the near future. So, a true digital 35-mm camera may not be too far away.

Conclusions

From all this, we can see that the digital scanning technology available to the consumer

at very competitive prices is just at the edge of the minimum resolution required to capture virtually all the image information in a negative (slide) or print. To really capture everything, perhaps an improvement of another factor of two or three in resolution is needed. But the resolution presently available is a reasonable compromise between capturing adequate image information and being able to conveniently store the resulting digital file with available storage technologies. Also, the printing technology is just at the limit of that needed to reproduce all the image information that can be visually discerned. So, in scanning and printing, the present technology is just at the limit of what is available in analog photography. There is still a little room for improvement, but the present technology is certainly capable of providing very impressive results.

But perhaps the most important feature of digital image storage is that the image never degrades. The contrast never diminishes, the colors never fade, and the image is always as pristine as when originally digitized. This is a very important feature in archiving images. Also, anyone may "print" digital images without a lot of costly equipment such as enlargers and lenses or peculiar chemicals, such as developers and fixers. All that is needed is a "photo-quality" color printer. Such a printer is likely a factor of five or 10 lower in price than a reasonable-quality color enlarger, without a lens. And, the cost of printing consumables will be much less than the cost of color-printing chemicals and color photographic paper. To complete the digital photography process, there is a wide range of digital image processing software available to provide anyone the ability to do very elaborate post processing of the digital image.

Digital cameras have a little farther to go to be truly equivalent to analog cameras. The standard analog camera format is the 35 mm. But even the high-end digital cameras are barely equivalent to a 16-mm format analog camera. The digital cameras are limited both by the size of available imaging electronics (CCD chips) and the memory limitations of the storage medium where the digitized image is stored. So, the digital-camera image size must be "improved" by about a factor of three in each dimension to be minimally equivalent to a 35-mm analog camera. We can be almost certain that it will happen, and very likely quite soon. ∎

REFERENCES
1. Eastman Kodak, EXR 200T 5293 Film Technical Data Sheet.
2. Eastman Kodak, Kodak Royal Gold 100 Film Technical Data Sheet.
3. AGFA-GEVAERT AG, AGFAColor Optima 100 Negative Film Technical Data Sheet.
4. John Russ, The Image Processing Handbook, 2nd Edition, CRC Press, Boca Raton, 1995.
5. J. Bernd, Practical Handbook on Image Processing for Scientific Applications, CRC Press, Boca Raton, 1997.
6. Mischa Schwartz, Information Transmission, Modulation and Noise, McGraw Hill, New York, 1970.
7. Eastman Kodak, PMI Student Filmmakers Handbook, 1997.
8. Hewlett Packard, Hewlett Packard PhotoSmart Scanner Product Data Sheet.
9. Nikon, Nikon CoolScan II Scanner Product Data Sheet.
10. Maxtor, Maxtor 5120 DiamondMax Plus High-Performance Disk Drive Product Data Sheet.

Table 2. Memory requirements for various scanner systems (excluding interpolation and without compression).

Scanner	Memory Required
HP Photo Smart	
2400 x 2400 dpi	???
Nikon CoolScan—35mm	
2700 x 2700 dpi	???
8.5 x 11 scanner	
1200 x 1200 dpi	404 MB
600 x 600 dpi	101 MB
300 x 300 dpi	25 MB
8.5 x 14 scanner	
1200 x 1200 dpi	514 MB
600 x 600 dpi	129 MB
300 x 300 dpi	32 MB
11 x 17 scanner	
1200 x 1200 dpi	808 MB
600 x 600 dpi	203 MB
300 x 300 dpi	50 MB

L.B. Cebik, W4RNL
1434 High Mesa Drive
Knoxville, Tennessee 37938
<cebik@utk.edu>

TRUE VERTICAL BEAMS FOR 2 METERS

Parasitic half-squares and bobtails

Sometimes a VHF operator needs a vertically polarized beam with a narrower beamwidth than a vertically oriented Yagi provides. Self-contained vertically polarized wire antennas popular on 160 through 40 meters offer a solution if scaled to VHF proportions and augmented by parasitical elements. The result of this exercise is a collection of reasonably small, relatively easy-to-build vertically polarized beams for 2-meters and up. If hams find enough uses, we shall just have to get used to talking about half-squares and bobtails at repeater club meetings.

Background: Beyond the vertical Yagi

The most common mono-directional antenna used on VHF is the Yagi tilted 90 degrees from its horizontal home position. Although this antenna performs well enough, its characteristics change significantly depending on orientation. **Figure 1** compares the horizontal and vertical patterns of a typical three-element 2-meter beam when oriented horizontally and vertically. (All azimuth patterns shown are for an elevation angle of 3 degrees for aluminum antennas at 30-foot heights over average ground conditions at 146 MHz.)

The gain of the vertical is down 1.5 dB from the horizontal due to the much wider bandwidth: 92 versus 60 degrees to the -3 dB points. Moreover, the vertically oriented beam has much less side rejection than its horizontal counterpart. Although these characteristics are useful for some purposes, for others, it would be helpful to have a vertically polarized beam with the pattern of the horizontal Yagi.

A side-fed quad suffers equally in terms of gain reduction and beamwidth. A square, two-element quad fed mid-side for vertical polarization has an excellent front-to-back ratio (>20 dB), but a gain of about 11.5 dBi and a beamwidth of 88 degrees or so. Like its vertically oriented Yagi cousin, the side-fed quad displays a broad forward lobe with poor side rejection.

In the Spring 1999 issue of *Communications Quarterly*, I suggested using a half-square as a bi-directional vertically polarized antenna.[1] Its figure-8 pattern had better gain than any other simple antenna, along with a set of very sharp and deep side nulls. The half square is a simplified version of the bobtail curtain, a three-element vertical antenna with a pair of half-wavelength phasing lines. **Figure 2** shows sketches of these two basic antennas, with some dimensions usable on 2-meters. **Figure 3** provides a comparison of the half square and the bobtail bi-directional patterns. Both patterns are broadside to the antennas.

The bobtail offers about 1 dB more gain, but in a much narrower beamwidth than the half-square: 48 to 77 degrees to the -3 dB points. Although the side nulls of the bobtail are shallower than those of the half-square, the overall narrower pattern of the bobtail can be very use-
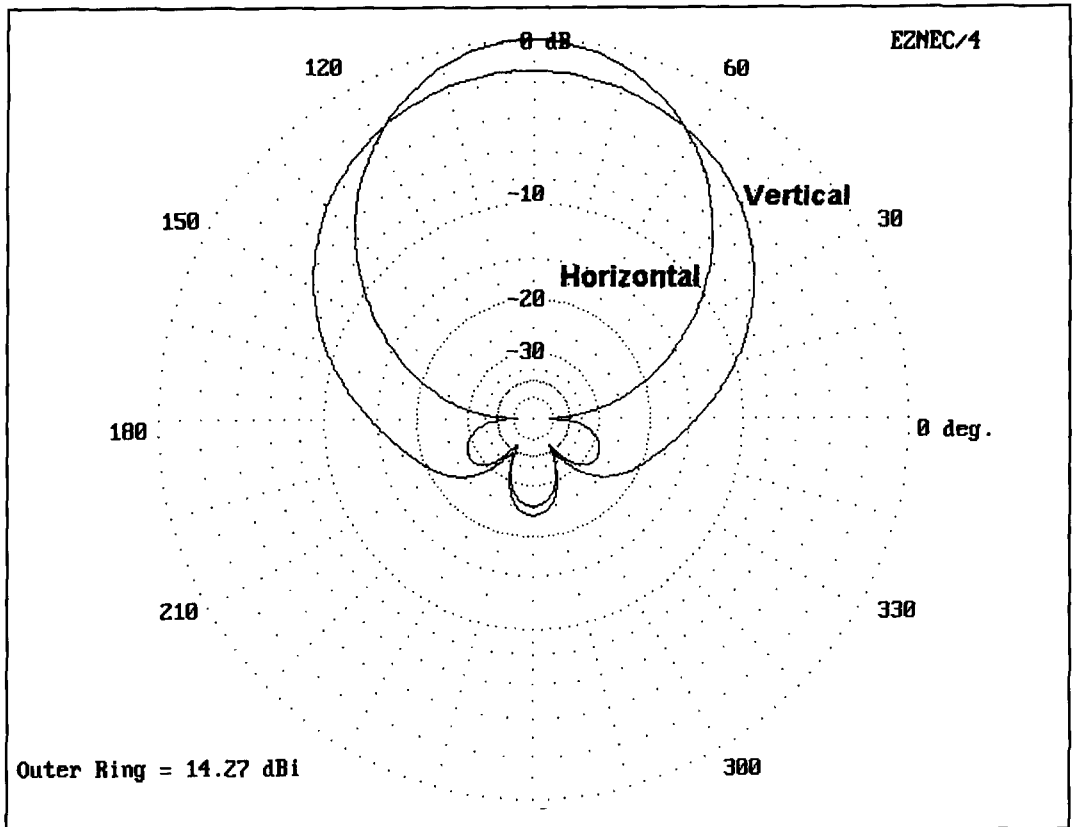
Figure 1. A comparison of far-field patterns of the same three-element Yagi oriented horizontally and vertically.
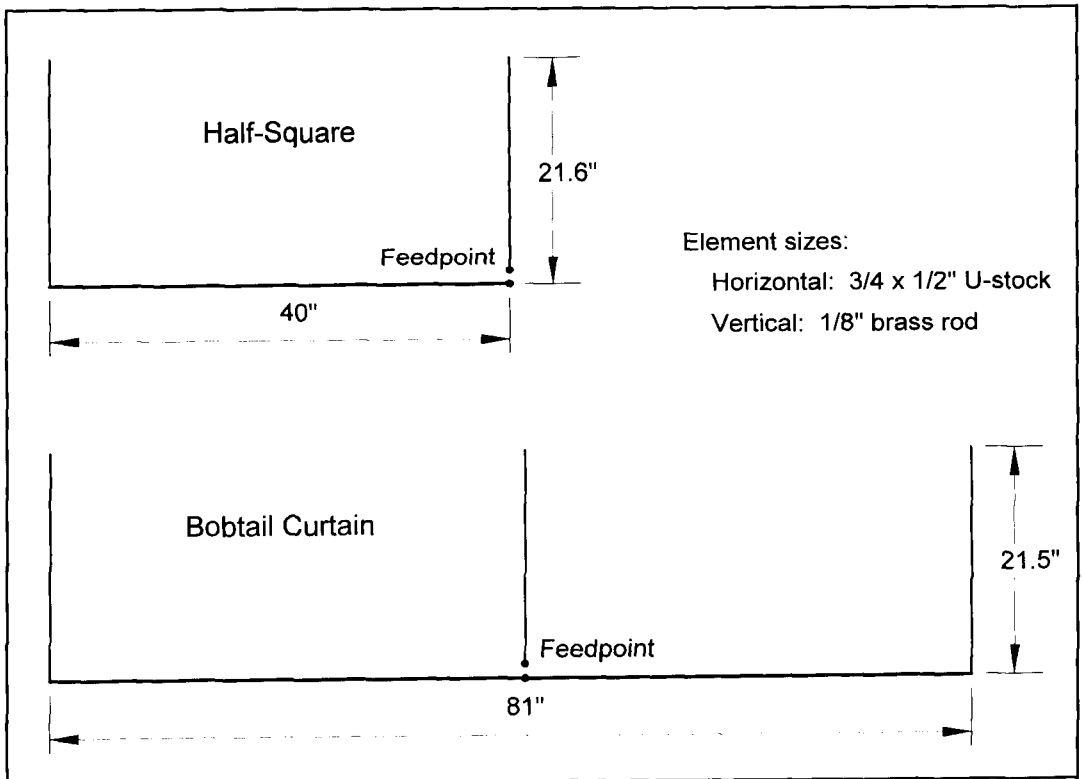

Figure 2. Dimensions of 2-meter half-square and bobtail antennas.

ful in reducing interference from quartering signals; that is, those only 45 degrees off the antenna's heading.

Both the half-square and the bobtail are interesting antennas, partly because there are numerous ways to look at them. First, they fall into the category of self-contained vertically polarized antennas that do not require the earth or a substitute earth structure to complete them. Indeed, one can look at the half-square as two bent dipoles joined at their horizontal ends and thus needing only one feedpoint. Second, one can view them as loops opened at a high-voltage point, and thus related to the side-fed deltas, squares, and rectangles. Changing the feedpoint of any of these antennas, especially moving that point to a horizontal wire, can change its operational characteristics.(One exception is that feeding the loose end of the half-square or the loose middle end of the bobtail with a high-impedance source will produce generally the same results as feeding the indicated junctions with a low-impedance source.) Third, one can view the antennas as phased quarter-wavelength verticals, where the currents in the horizontal wire are equal and opposite at either side of center, thus canceling to a large measure any horizontally polarized radiation fields.

The currents in the vertical legs of the half-square are virtually the same and out of phase, resulting in their adding broadside to the plane of the antenna. (Since the voltages are also out of phase on the two verticals, the net effect is the same as two independently fed verticals fed in phase.) Because the feedpoint is offset to one end, the currents have enough differential due to wire losses in the horizontal section to offset the pattern by about 4 degrees from true bidirectionality. The center feedpoint of the bobtail places equal current on the end verticals, producing a true 180-degree pattern. You can think of the center bobtail conductor as two wires in parallel, which yields a 1:2:1 current ratio (minus wire losses) among vertical elements, counting from one end to the other.

The disadvantage of the bobtail is that it is close to a full wavelength long in thin wire versions. At 2-meters, it is a little under 7 feet long —and a little strange looking to those unfamiliar with the antenna in its pointing-up position. (Most lower HF versions point down for reasons related to physical construction and to elevation of the maximum current points.) Nonetheless, 7 feet is a short span in overall antenna terms and easily supported. The 40-inch-long half-square is even easier to support and lends itself to almost endless construction variations.
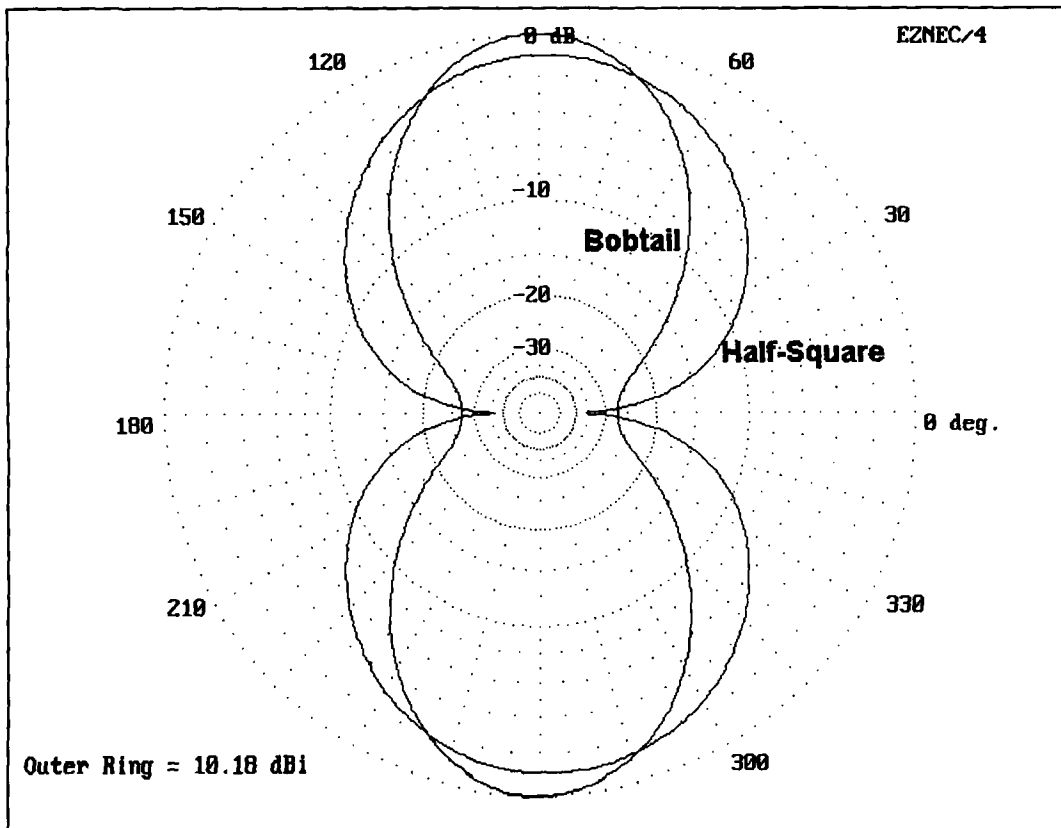

Figure 3. A comparison of the azimuth patterns of 2-meter versions of the half-square and the bobtail curtain.
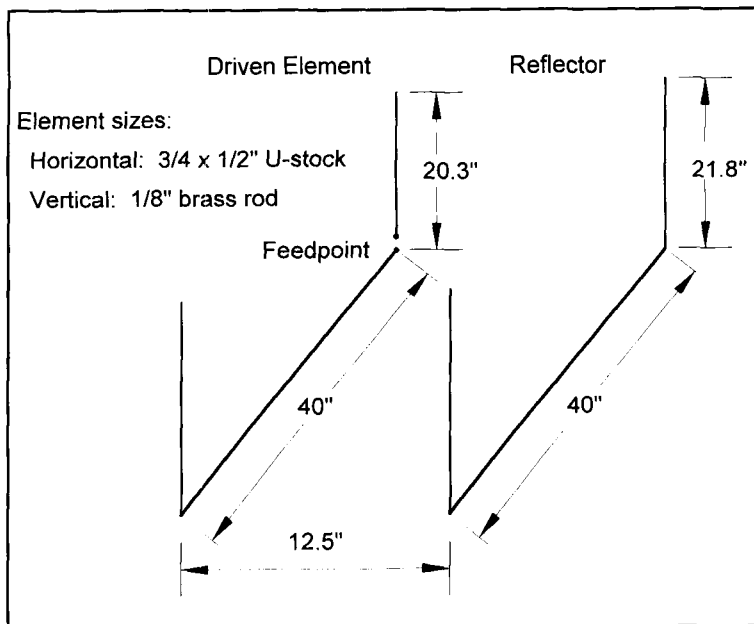
Figure 4. Dimensions of a 2-meter, two-element half-square beam.

However, before leaping into construction, let's think about squeezing more performance out of these antennas.

Parasitic half-square and bobtail beams

When it comes to "square" antennas, two halves do not make a whole. Rather, they make an interesting parasitical beam with a driven element and a reflector, as sketched in **Figure 4**. (The reason I do not call this whole is that I shall add another element in a moment.) The result is a vertically polarized beam with nearly 13 dBi gain at the design center frequency, over 18 dB front-to-back ratio, a feedpoint impedance compatible with coaxial cable, and a pattern very similar to that of a horizontal Yagi. The side nulls are very deep, while the beamwidth is about 70 degrees to the -3-dB points. The entire rear quadrant is flat and better than 17 dB down. **Figure 5** shows the azimuth pattern of this antenna.

Because the horizontally polarized radiation is not completely suppressed, it contributes to the overall pattern by slightly widening the beamwidth and tilting the forward lobe about 3 degrees from true. In some applications, cross polarization of the desired signal from the horizontal component of the antenna pattern may narrow the effective beamwidth of the antenna still further. Compare the "two-element" half-square beam pattern to that of the vertically polarized Yagi to get a sense of the differences to be expected in performance.

One can achieve about 1.5 dB more gain, 2 dB more front-to-back ratio, and a beamwidth of 60 degrees through the addition of one more element, as shown in **Figure 6**. With the dimensions shown, the antenna can be fed with coax from about 145.5 to 148 MHz with under 2:1 SWR. **Figure 7** shows the azimuth pattern of the beam at design center frequency. Gain is fairly constant across this bandwidth, but the front-to-back ratio slopes from its high at the design center frequency to about 12.5 dB at the band edges. With a different selection of dimensions, one can achieve about 0.3 dB additional gain and about 2 dB additional front-to-back ratio, but at a cost of a much lower feedpoint impedance and a much narrower bandwidth. The dimensions I settled on appear sensible for the portion of 2 meters used for FM and other vertically polarized services.

Both the two- and three-element half-squares provide vertically polarized patterns that are otherwise comparable to the horizontally polarized patterns of horizontally oriented Yagis. There may be occasions calling for even narrower vertically polarized beamwidths, perhaps in the neighborhood of 45 degrees or so. For this purpose, a two-element bobtail may prove useful. Although seemingly more ungainly, the two-element bobtail has only six vertical spikes—the same number as the three-element half-square.

Figure 8 sketches the elements of a two-element parasitical bobtail, with the dimensions used in the test antenna. Modeled performance predictions suggest a gain of about 13.4 dBi (halfway between the two- and three-element half-squares). The front-to-back ratio is about 18 dB, with the entire rear quadrant at this figure or better. **Figure 9**, the azimuth pattern for the antenna, shows the extreme elongation of the bobtail pattern. Feedpoint impedance is in the 50-ohm range, suitable for coaxial feedlines. The pattern holds up reasonably well across 2 meters, with better than 12.8 dBi gain and at least 11 dB front-to-back ratio across the span from 144.5 to 148 MHz. SWR is less than 2:1 across the whole band.

A three-element bobtail is certainly possible. However, its 9-spike ungainliness and the likely narrowband response put it on the back burner for this exercise.

Building and testing multi-element half-squares and bobtails

Constructing test half-square and bobtail beams is strictly a hardware depot exercise. The following are the sum of materials used: 3/4- by 1/2-inch channel (U-shaped) aluminum stock, 3/4-inch hardware depot aluminum tub-
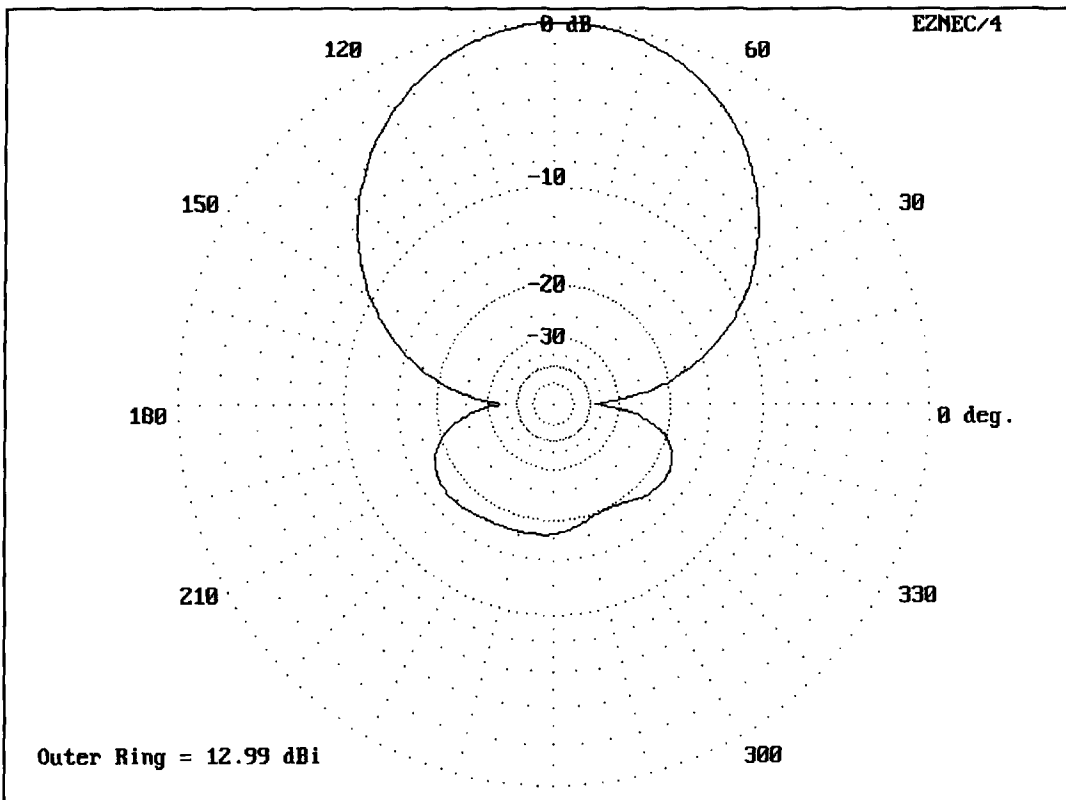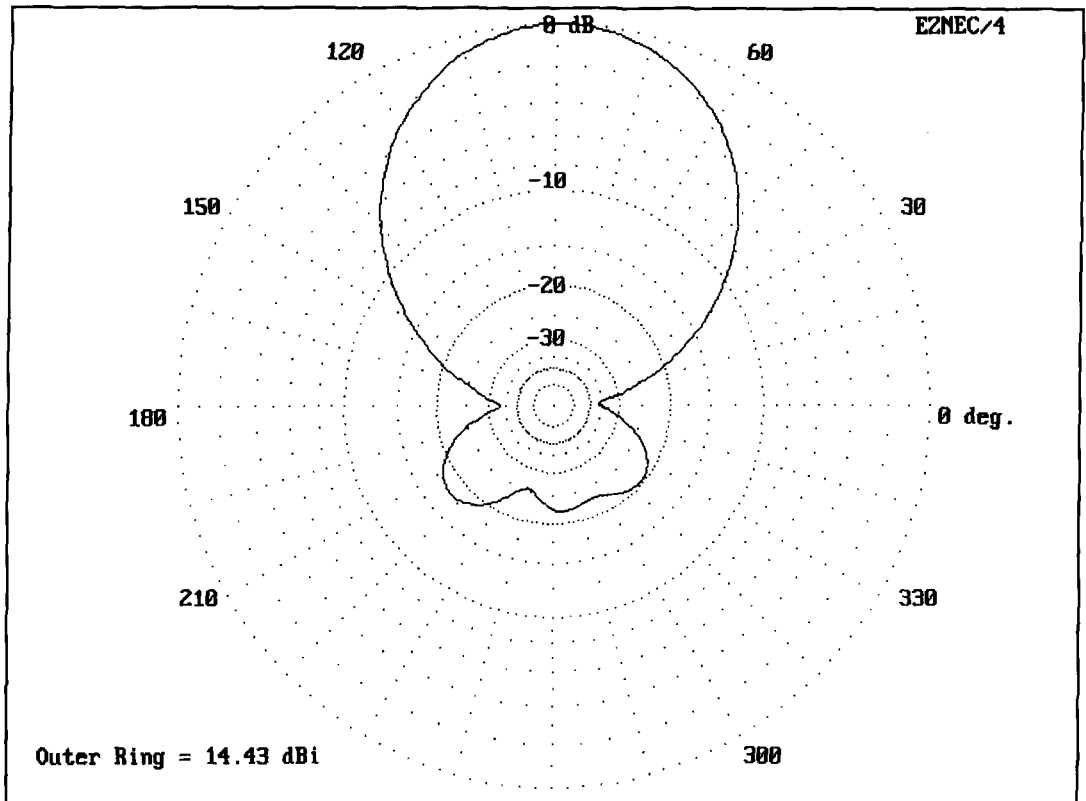
Figure 5. The 3-degree-elevation azimuth pattern of a two-element half-square beam (aluminum elements, 30-foot height, average ground).

ing, 1/8-inch brass rods, 1/4-inch aluminum rods, female coax connectors, 1/2-inch nominal Schedule 40 PVC, miscellaneous PVC couplings, and some nuts and lock washers to fit the threaded ends of the rods. All dimensions are as shown in the outline drawings of the various configurations. **Figure 10** shows the basics of the element construction for the test models.

For one prototype, a length of 3/4- by 1/2-inch channel or U-stock forms the horizontal portion of either the half-square or the bobtail. All vertical elements are 1/8-inch diameter
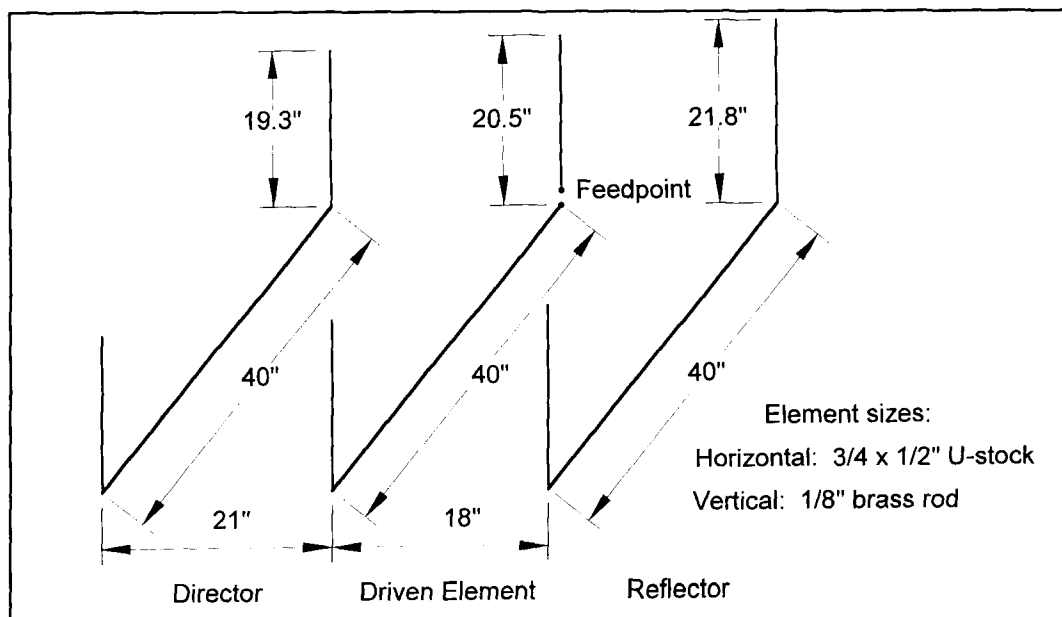
Figure 6. Dimensions of a 2-meter, three-element half-square beam.

Figure 7. The 3-degree-elevation azimuth pattern of a three-element, half-square beam (aluminum elements, 30-foot height, average ground).

brass rod. The rod is threaded at one end and secured through a hole in the channel with top and bottom lock washers and nuts. I used an antioxidant at these bi-metal junctions.

The "driven" element consists of two parts. One is a flangeless SO-239 coax connector mounted in a 5/8-inch hole in the U-stock. The other part is the brass rod, ground to fit the solder cup of the connector. I applied solder to both items before joining them. A PL-258 screwed firmly into the coax connector holds the solder cup rigid for this operation.

For other prototypes using 3/4-inch diameter aluminum tubing, I used U-stock for the PL-258 simply by "grafting" a short length onto the tubing. For half-square models, this graft forms an end piece. Bobtails require splitting the horizontal tubing in the middle and splicing with U-stock and the coax connector. Because the PVC mounting system uses off-center support, the splice undergoes no significant mechanical stress.

Support for the antenna elements is shown in **Figure 11**. Essentially, the overall antenna supports are exercises in PVC welding/gluing. A combination of Tees, 90-degree elbows, and straight lengths of 1/2-inch nominal Schedule 40 PVC combine to form a sturdy, fairly light support structure. The only difference between using U-stock and tubing is the orientation of

the upper Tees. They are perpendicular to the horizontal antenna element with U-stock, and the element rests in grooves cut in the Tees. For aluminum tubing, the Tees align with the element, which passes through them.

A single element can be supported with three Tees, two elbows, and about 12 to 24 inches total of PVC tubing (not counting the vertical section), depending upon whether a half-square or a bobtail is in question. The bobtail feedline will come down next to the mast. However, the line to the half-square proceeds to the side, and its weight should be accounted for in balancing the elements. A multi-element support requires more Tees, elbows, and straight lengths. The key to the multi-element support is to be sure that the center-to-center distance between element supports is exactly the desired spacing between elements.

I attached the U-stock elements to the supports by first sawing two 1/2-inch deep and 3/4-inch apart channels in the Tees that fit under the U-stock of the horizontal portion of the elements. The channels are just wider than the U-stock thickness—about 3/16 inch for the 1/8-inch thick stock I used. The completed element is balanced in the two Tees supporting it. I have used two methods to fasten the U-stock to the Tees. As shown in **Figure 12**, one method is to use a stainless steel-sheet metal
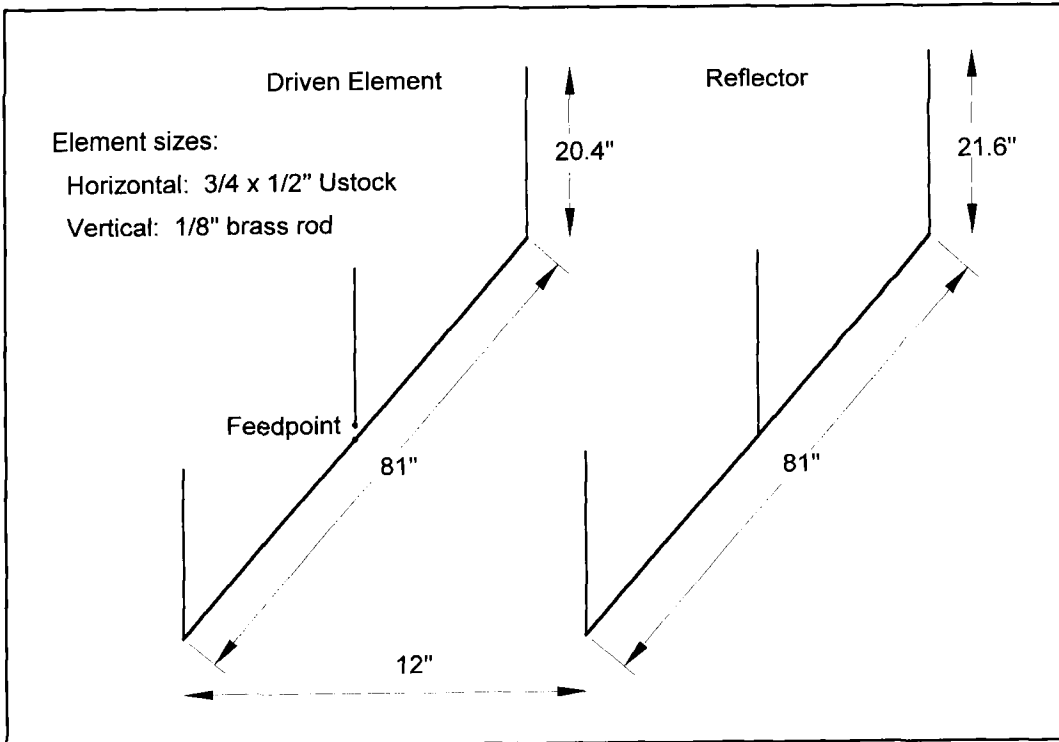
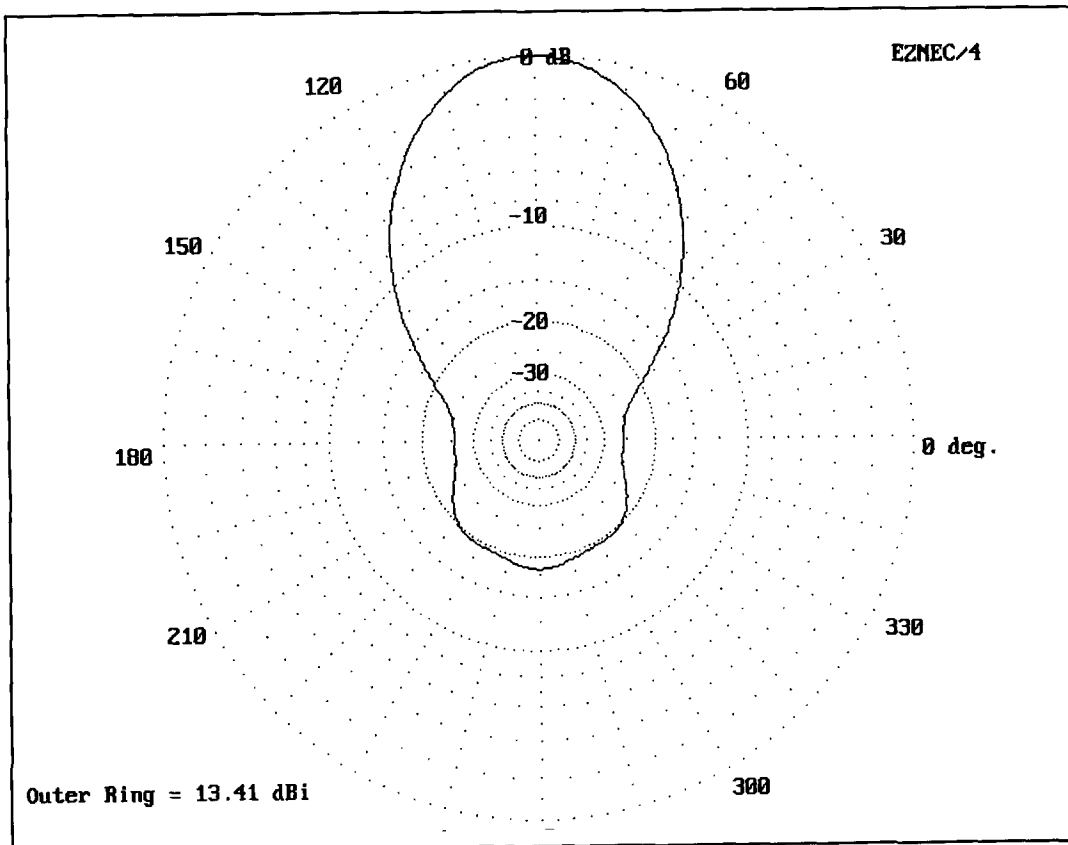Figure 8. Dimensions of a 2-meter, two-element bobtail beam.

Figure 9. The 3-degree-elevation azimuth pattern of a two-element bobtail beam (aluminum elements, 30-foot height, average ground).
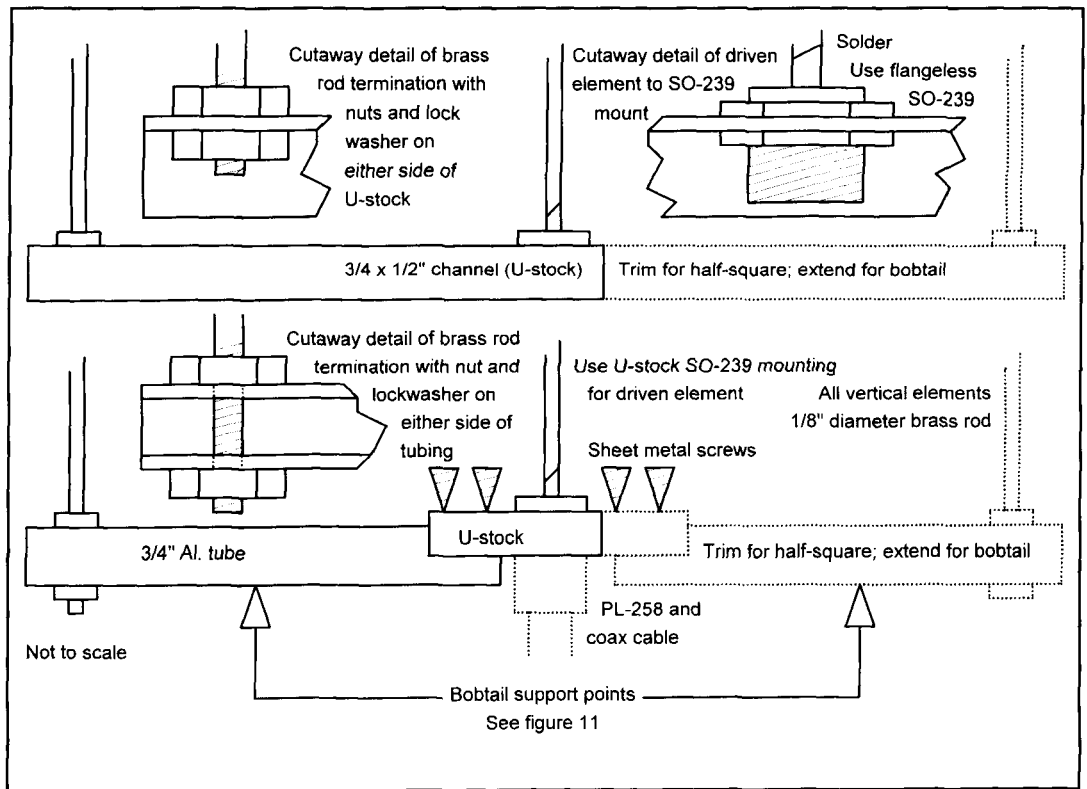
Figure 10. Half-square and bobtail element construction detail.
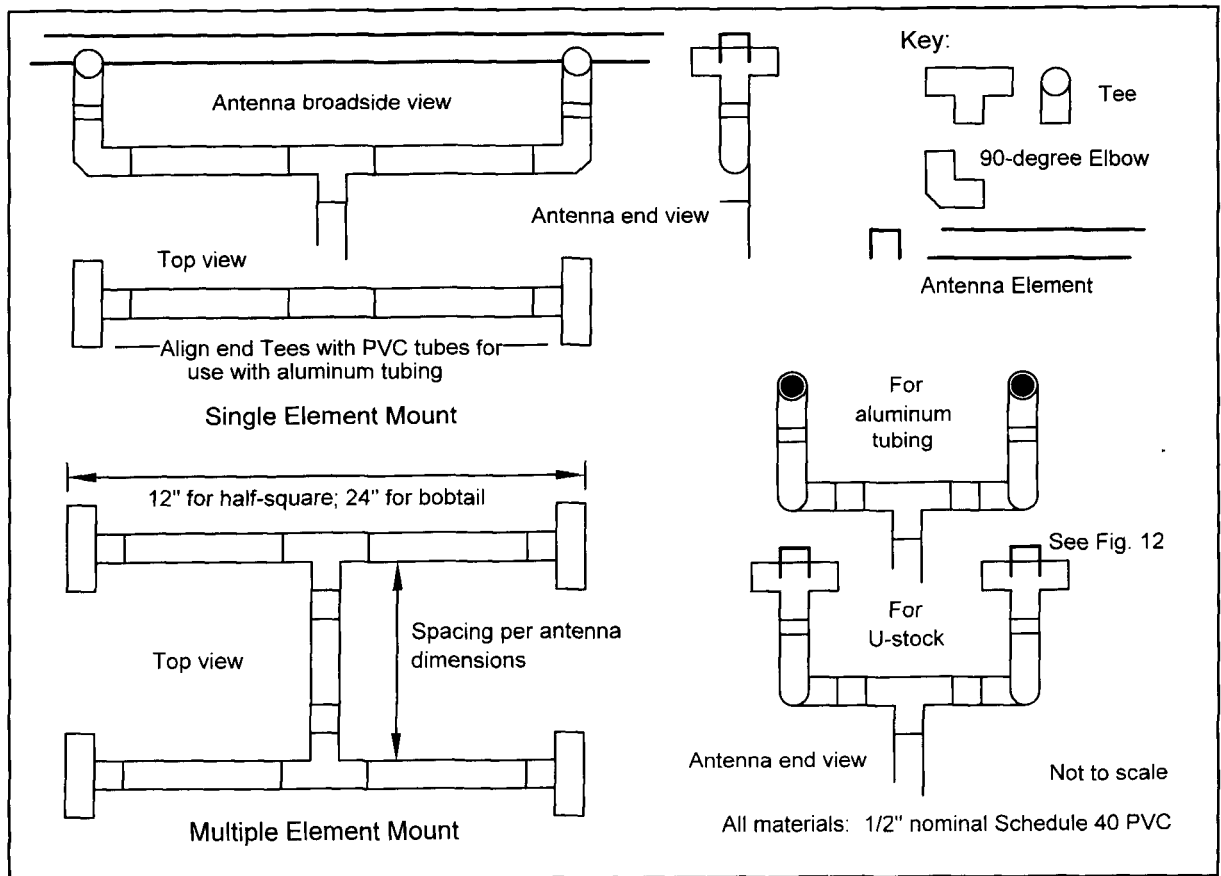

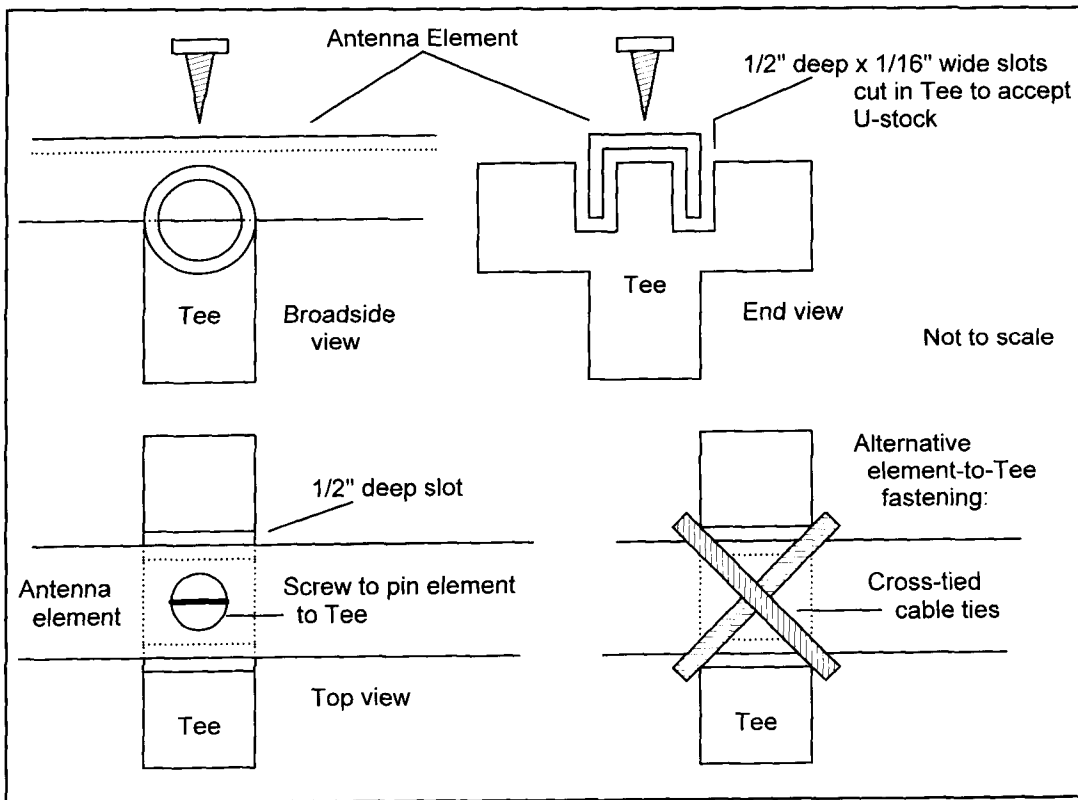Figure 11. General outlines of half-square and bobtail beam mountings.

Figure 12. Details of antenna-to-Tee attachment.

screw through the U-stock into the Tee. For some tests, I simply lashed long cable ties in an X formation across the junction. Aluminum tubing versions of the antennas simply use a sheet-metal screw through each support Tee into the horizontal element to lock it in place. Both systems have withstood significant winds with no movement of the antenna pieces.

The weak link in this construction is support of the driven element brass rod in its SO-239 base. The short solder joint is not sufficient to support the rod without further bracing. To accomplish this, I have used a variety of "caps" constructed from scrap PVC. **Figure 13** shows a couple of versions. One used a short length of 3/4-inch nominal PVC with a channel cut out to press fit over the U-stock. A PVC end cap with a hole in the top just able to slip the brass rod is glued to the tube. Silicone sealant secures the junctions of the tube and the channel, as well as the point where the rod emerges from the cap. This keeps out rain. However, bottom openings are left open to prevent moisture build-up inside this fixture.

A second method uses a short section of 1-inch nominal PVC with a section removed lengthwise to press-fit over the U-stock. A hole in the side passes the brass rod. After struggling to fit some sort of caps on the ends for rain protection, I found it easiest to built up silicone sealant to plug the ends. I also sealed the rod-

tube junction and created two weep holes along the edge.

I have used the cap and the tube on versions of the half-square and the bobtail. Both have proven to be good braces for the driven element rod, with no mechanical problems so far. Nonetheless, I would recommend a superior element bracing system for permanent installations.

Testing of the various beams was limited by my test range: a ladder-supported length of TV mast and a few pieces of test equipment. Tests with single-element antennas showed that the modeled dimensions were very close to reality—within about 100 kHz at 146 MHz, despite the use of U-stock instead of a round element. Therefore, for the multi-element models, reflector and director elements were precut, and the antenna pruned to SWR minimum at 146 MHz.

Although I lack the equipment and range to quantify the test results, indications point to performance fairly close to modeling predictions. The half-square beams showed excellent front-to-back characteristics, dropping repeaters into a noisy condition off the rear, and even more so off the sides. The narrower bandwidth of both the single-element and the multi-element bobtail was very evident. Also evident with the two-element bobtail was the smooth rear quadrant pattern of rejection, with the signal rising (and then falling) quite abruptly as I hand-turned the mast.
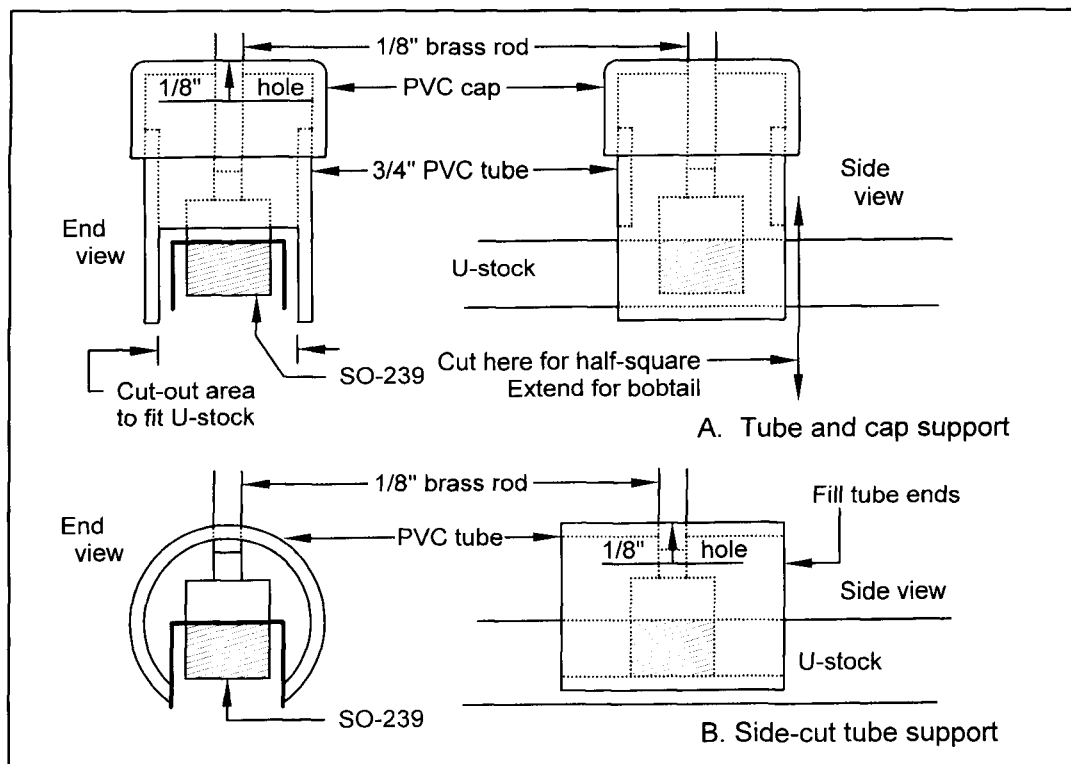
Figure 13. Details of two driven element supports.

Although the U-stock has worked very well with the test models, I would recommend the use of round stock for the horizontal portions of more permanent installations, especially where the antennas are likely to suffer heavy winds. U-stock forms an air foil that may stress the assembly in unpredictable ways. Under certain wind conditions, U-stock can create vibrating twists that wobble the elements in and out of alignment. The vibration can be damped simply by running screws and nuts across the sides about two-thirds outward from center. Alternatively, one might use 1-inch square stock from the hardware depot. For home installations up to about 30 feet or so, channel or square stock may prove easier to work with in the long run than tubing.

Conclusion

The half-square and bobtail beams modeled and tested at this location may have a niche among amateur VHF antennas. As predominantly vertically polarized beams with narrower beamwidths than Yagis and quads turned on their sides, the half-square and bobtail beams offer good gain and front-to-back characteristics that may be useful in FM, repeater, packet, and other services using vertical polarization. Construction is very straightforward and suited to the home shop.

Perhaps the one drawback of these antennas is their appearance. My YL, N4TZP, has

described them as resembling a porcupine with a punk hair-do. If you are not put off by unfamiliar appearances, the beauty of these antennas may show up in their performance.

Should you wish to do further research into these interesting antennas, you will find a host of resources in the literature. See the bibliography on half square (Notes 1-4) and the bobtail curtain (Notes 5-10).

Acknowledgments

Notes
1. Ben Vester, K3BC, "The Half-Square Antenna," *QST*, March 1974, pages 11-14.
2. Robert Shiers, N0AN, "The Half-Square Antenna," *Ham Radio*, December 1981, 48-50.
3. Paul Carr, N4PC, "A Two-Band (12 & 17 Meter) Half Square with Coaxial Feed," *CQ*, September 1992, pages 40-45.
4. Paul Carr, N4PC, "A DX Antenna for 40 Meters, *CQ*, September 1994, pages 40-43.
5. Woodrow W. Smith, W6BCX, "The 'Bi-Square' Directable Array," *Radio*, April 1938, pages 36-38, 72.
6. Woodrow W. Smith, W6BCX, "Bet My Money on a Bobtail Beam," *CQ*, March 1948, pages 21-23, 92-95.
7. Woodrow W. Smith, W6BCX, "The Bobtail and Inverted Ground Plane," *Ham Radio*, February 1983, pages 82-86.
8. Woodrow W. Smith, W6BCX, "Bobtail Curtain Follow-Up: Practical DX Signal Gain," *Ham Radio*, March 1983, pages 28-30.
9. Jim Gray, W1XU, "Those Amazing Bobtails, 73, February 1982, pages 30-32.
10. Bud Hippisley, K2KIR, "The Upside-Down Vertical," *The DX Magazine*, August 1989, pages 8-10.

REFERENCE
1. L.B. Cebik, W4RNL, "A Bidirectional Vertical Antenna," *Communications Quarterly*, Spring, 1999, pages 65-70.

Ian Poole, *G3YWX*
5 Meadway, Staines
Middlesex, TW18 2PW
England
<http://website.lineone.net/~ian_poole>

QUARTZ CRYSTALS

A resumé

Quartz crystals are widely used in the electronics industry, particularly in RF circuits. Here these crystals provide a level of performance that can't be matched by other components. While crystal technology may appear to be as old as the hills, nothing could be further from the truth.

Crystals are used as the form of a very high-performance resonant circuit. Their chief advantage is the exceedingly high levels of Q that can be reached. Values between 10,000 and 1,000,000 can be obtained, depending upon the specification required. These Q levels mean that crystals can be used in a number of areas. Naturally, they can be used in filters to provide levels of performance not achievable by other means. They can also be used in oscillators. The Q values that can be obtained result in very high levels of stability and exceedingly low levels of phase noise, an essential factor when used for a reference oscillator for a synthesizer.

Unfortunately, these dependable circuit components are often taken for granted. The low cost for their high performance and their availability mean we don't need to ask many questions about them. Little is seen in the literature about what they are and how they operate. Often a bit of background information is interesting and informative, helping us to make the most of the crystal's exceptional qualities.

What is quartz?

Quartz is a unique material among the many thousands of minerals found on the Earth's sur-
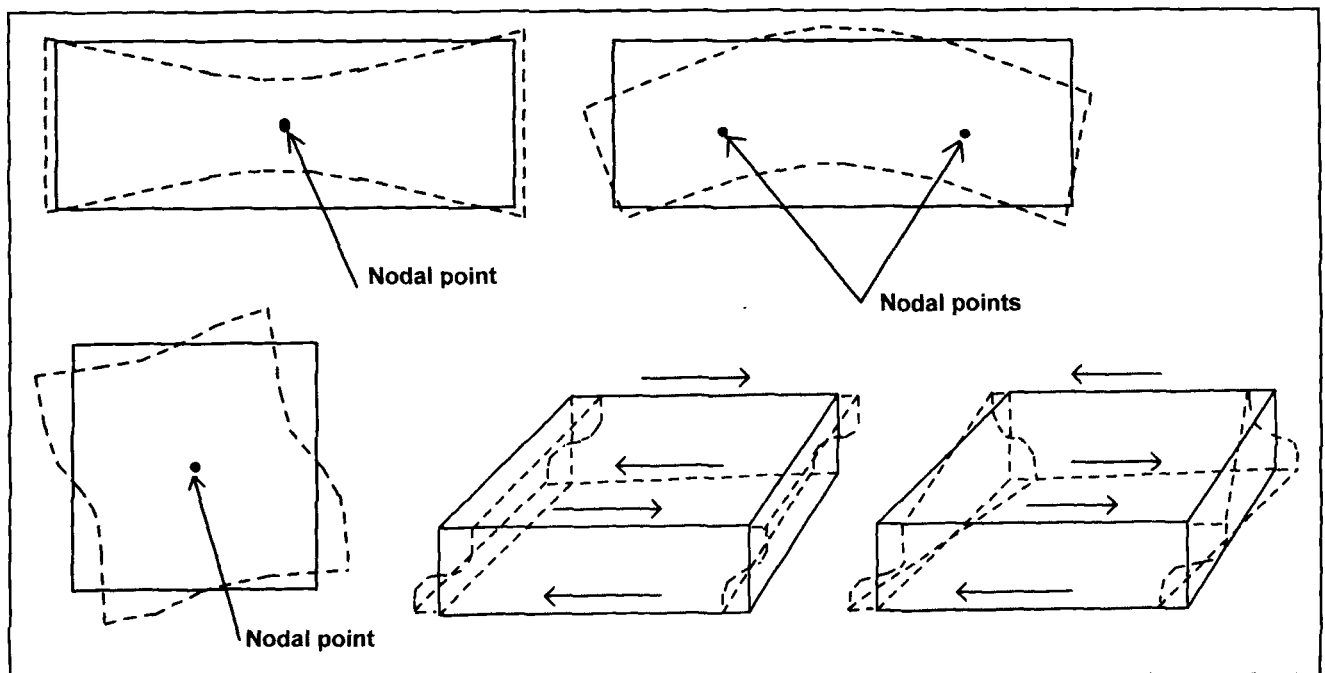
Figure 1. Modes of vibration of a crystal.
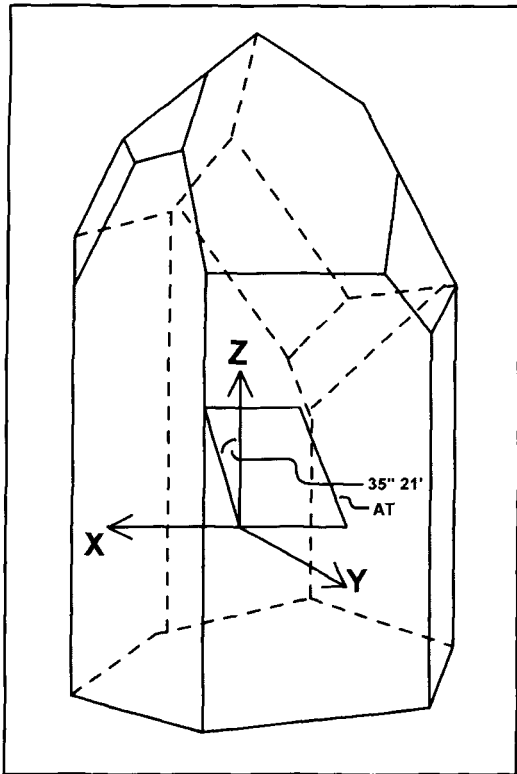
Figure 2. Crystal structure.

face, and it possesses a rare combination of properties that enable it to be used in electronic components. Quartz is a crystalline form of the mineral silicon dioxide—the most abundant mineral on the Earth. It is distributed all over the world and is the major constituent of many rocks, such as sandstone. It is sometimes in the form of pure deposits. Quartz deposits in rock fissures contain valuable minerals. Precious metals, including gold, are often found in sufficient quantities to warrant the mining of quartz to recover the gold or other precious metals. Quartz is also the primary constituent of sand, and this alone demonstrates its abundance.

In its pure state quartz is white, but when impurities exist within the mineral it takes a variety of colors which makes it popular for jewelry. Quartz can be carved into many shapes, and this property is exploited by jewelry designers who create some outstandingly beautiful objects d'art. Some specimens have a very shiny surface, which make them particularly impressive for jewelry.

Quartz crystallizes into a rhomboidal shape. The crystals can vary enormously in size from small particles to crystals of around a metric ton. The mineral is also very hard, having a hardness of seven on the Mohs hardness scale. It also has a specific gravity of 2.65.

Quartz exhibits a piezoelectric effect, which I will describe later. This is the reason for its

widespread use within the electronics industry. Interestingly, the material also has an optical property that enables it to rotate the plane of polarized light. Accordingly, it is used in polarizing microscopes.

Manufacture

Naturally occurring quartz is not suitable for electronic use as it contains too many defects and impurities in its crystal structure. As a result, all the quartz used for electronic components is manufactured synthetically.

The process starts by dissolving powdered natural quartz in water at 400 degrees centigrade under very high pressure. Small seed crystals are then suspended in the solution, and the temperature of the solution is slowly reduced so it can crystallize out onto the seed crystals. Using this method, crystals of several inches in length can be grown.

Once the basic crystals have been obtained, it is possible to make the individual components. Blanks are first cut from the main crystal using diamond wheels. These must be used due to the extreme hardness of quartz. The angle of the cut relative to the axes of the crystal is also very important, as this determines many of the properties of the crystal.

After cutting, the blank is lapped to the approximate thickness for the frequency required. This is done using either silicon carbide or aluminum oxide paste, once again because of the hardness. The paste is very fine, so an extremely good finish can be obtained. Even so, chemical etching is used to provide near perfect results.

The next stage in the manufacturing process is the addition of the electrodes. These are normally gold or silver and deposited onto the faces of the crystal. The thickness can be controlled to very fine limits, and the crystal must be trimmed to its final frequency. Once the contacts are added, the crystal is mounted mechanically in such a way that the mechanical losses are reduced. Finally, the crystal is placed into its can, which is either evacuated or filled with an inert gas.

How they work

The operation of quartz crystals is based around the piezoelectric effect mentioned earlier. Pierre Curie and his brother Jacques first discovered this effect in 1880. They noticed it in a number of materials including quartz and Rochelle salt. They named the effect piezoelectricity after the Greek word "piezein" meaning to press.

The piezoelectric effect is an unusual phenomenon where a mechanical stress applied to

a material results in a potential being generated within it. Conversely, when a potential is applied across the material, a stress is set up within it. This is caused by the displacement of ions in crystals that have a non-symmetrical unit cell. When the crystal structure is stressed, the ions in each unit cell are displaced causing electric polarization of the unit cell. Because the crystalline structure is repeated, the effect accumulates over the entire crystal and causes an electric potential to appear across certain of its faces.

Similarly, when an electric potential is placed across certain faces of the crystal, the reverse effect takes place and the crystal is distorted. Accordingly, it can be observed that the piezoelectric effect operates in both directions, changing electric potentials into movement and deforming the crystal into an electric potential. As a result, materials exhibiting this effect are used in a variety of applications from microphones to movement sensors and, in the case of quartz, as the resonant element in electronic circuits.

Looking more specifically at the way a quartz crystal is used, it is possible to place an alternating waveform across the crystal and cause it to vibrate. Alternatively, if a vibration is made to bend and flex a crystal, then an electrical waveform corresponding to the mechanical movements can be seen. In essence, an incoming signal is converted into mechanical vibrations by the piezoelectric effect. The mechanical resonances of the crystal then affect the signal that is then reconverted back into an electrical signal. The overall effect is that the electrical signal is affected by the mechanical resonances of the crystal, which offer very high levels of selectivity.

A crystal can vibrate in a number of ways and each mode has its own resonant frequency. The actual situation is quite complicated because the crystal has different elastic constants in different directions and this alters the resonant properties for that orientation. However, there are three basic vibrations of crystals: longitudinal, flexural, and shear. Longitudinal vibrations produce a displacement in the direction of the length as shown in **Figure 1A**. In the flexural mode, the crystal bends or flexes as shown in **Figure 1B**. There are two types of vibrations within the shearing mode. The first is a face shearing motion like that shown in **Figure 1C**. This is normally associated with low frequency vibrations. The second type is a high frequency thickness shear mode that is shown in **Figure 1D** and of which there are two slightly different types.

Associated with each of these modes are overtone modes that may be exploited, providing still further resonances in the crystal. The Q
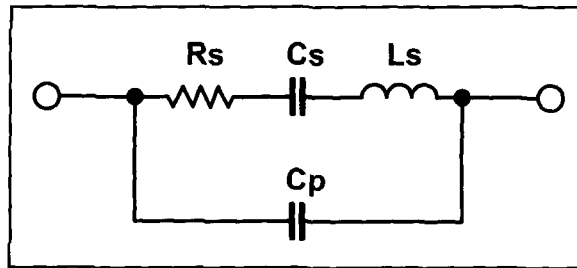
Figure 3. The equivalent circuit for a crystal.
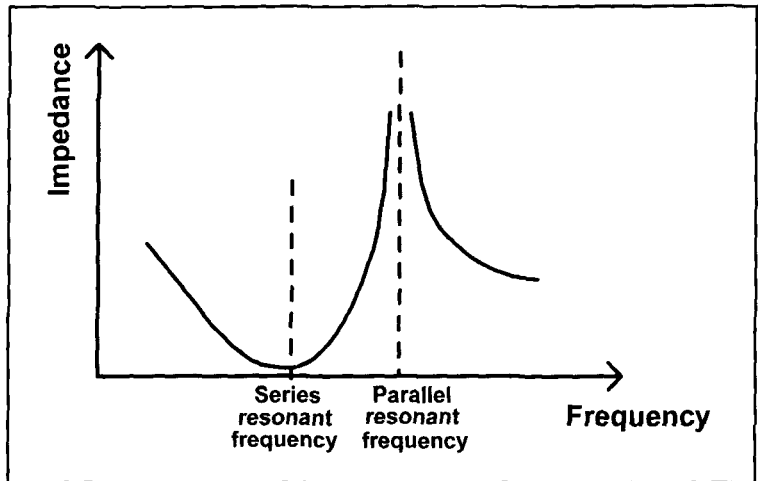
Figure 4. Impedance variations of a crystal.

for these overtone modes is normally of the same order as that of the fundamental mode; but the activity is normally less, reducing with higher orders of overtone. As a result, overtone modes are usually restricted to the lower values. Three and five are the most common.

Crystal cuts

The angle at which the crystal blank is cut relative to the axes of the main crystal is very important. It governs many of the properties of the crystal from the way in which it vibrates to other properties like the temperature stability. It's very important to choose the correct cut for any application. Remember that the elastic constants are different in different directions. The piezoelectric effect also changes according to the plane. Any spurious responses can be minimized by choosing the correct cut, and the possibility of unwanted signals can be reduced. However, it's always best to incorporate a tuned circuit into the oscillator to ensure that the correct mode is excited. There have been instances where oscillators have started with the crystal in the wrong mode and, as a result, they have generated a signal on the completely wrong frequency.
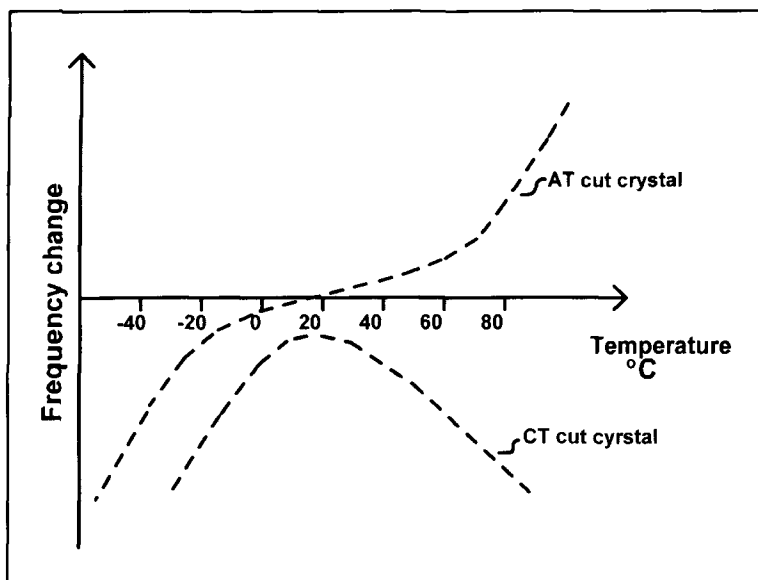
Figure 5. The variation of temperature coefficient with temperature.

25 kHz higher than the equivalent harmonic of the fundamental.

Equivalent circuits

To explain how a component operates in an electrical circuit, it's often convenient to describe it in terms of basic resistors, capacitors, and inductors. By doing so it is possible to see how the crystal will function in a circuit and to gain more of an idea of how it operates.

Figure 3 shows the equivalent circuit for a crystal consisting of four components. Cs and Ls represent the motional capacitance and inductance while Rs represents the frictional losses. Cp results from the capacitance of the plates and leads in the component. There are two ways in which the crystal can resonate. The first is in a series resonant mode where the combinations of Ls, Cs, and Rs govern the characteristics. Being a series resonant mode, the impedance of the circuit falls to a value governed by the series resistance, Rs. In this mode, the external circuit has very little effect on the resonant frequency.

There is a second resonant frequency where the combination Ls, Cs, and Rs have an inductive reactance equivalent to the parallel capacitance Cp. This is a parallel resonant frequency where the impedance rises to a maximum value at resonance. Unlike the series resonant mode, this mode is dependent upon the external circuit for its exact resonant frequency. As a result, crystals for use in this mode have an external load capacitance specified. Normally, it is either 30 pF or sometimes 20 pF for higher frequency crystals. Circuits for use with a crystal in this mode should be designed to present the correct load capacitance if the crystal is to operate exactly on its intended frequency. Variations in capacitance will provide small but important variations in frequency.

The difference between the two resonant points is generally quite small, usually less than about 1 percent. This means that there is an enormous variation in impedance over a small frequency difference where the crystal changes from a low-impedance series resonant mode to a high-impedance parallel resonant mode. This provides an impedance plot as shown in **Figure 4**.

Temperature stability

The temperature stability of crystals is very important. Often they are used as the resonant element in accurate or close-tolerance frequency sources. For example, they may be used as the clock oscillator in a frequency counter. Frequency accuracy is also important when

A further factor is dependent upon the plane, and this is the change in resonant frequency with respect to temperature. Naturally, this factor is particularly important when using a crystal as a frequency reference.

For most crystals operating above about 500 kHz, a cut known as the AT cut is used. This cut is made at an angle of 35 degrees 21 minutes to the axis running the length of the crystal shown in **Figure 2**. The AT cut provides a high level of activity combined with a low level of spurious responses and a good temperature stability. For lower frequency applications, cuts referred to as CT or DT may be used; for the very low frequency ones, the NT is chosen. These all have their particular orientations within the crystal and will be cut to meet their specific requirements.

Overtones

It is possible to manufacture crystals to operate at frequencies up to about 30 MHz. Beyond this they become very thin, fragile, and difficult to construct. To overcome this problem and obtain crystals that can oscillate at frequencies well above 100 MHz, they can be run at odd overtones of the fundamental frequency. To achieve this, the crystal is specially manufactured to optimize its operation at the overtone, and the oscillator circuit must be tuned to the correct frequency to ensure that the correct overtone is selected. When operating in this mode, it must be remembered that the overtone frequency is not an exact multiple of the fundamental. Owing to the way the crystal operates, the overtone frequency is approximately

crystals are used in an oscillator for the reference in radio receiver where the tuning accuracy is dependent upon the tolerance of the reference. In applications like these, it's necessary to ensure that the effects of temperature variations are as small as possible.

The temperature coefficient of a crystal is dependent upon the cut. **Figure 5** shows that the coefficient varies with temperature, and, for the AT cut, the minimum variation is reached just above 0 degrees C—quite acceptable for most applications between 0 and 70 degrees C.

Crystal ovens

In many applications, a crystal oscillator on its own is quite adequate. However, in some applications, higher degrees of accuracy are needed. One common example is the clock oscillator in a timer or frequency counter. Here the accuracy of the entire instrument is dependent upon the accuracy of the oscillator.

To achieve the best accuracy, crystal oscillators are often contained in what is called an oven. Here the whole oscillator circuit is placed in a small temperature-controlled environment. These units often cost around $150 and contain a crystal oscillator that is normally run from an internal regulated supply. The oscillator is then mounted in some thermal insulation which is often a form of foam. This is contained in a metal can and usually measures about two inches cubed.

The temperature is maintained at an elevated level, above the maximum external operating temperature, and is held constant using a small heating element and a thermostat. By removing any temperature variations, the accuracy of the oscillator can be improved by about a factor of 10 over a normal oscillator. When the accuracy of an instrument depends on the oscillator, the added cost can be worthwhile.

Aging

Like many other components, crystals change with time. As they are usually designed to be an accurate element in a circuit, especially when used in measuring instruments, this factor can be of great importance as the resonant frequency changes as the crystal ages.

One of the main reasons for this change arises from disruptions in the crystal lattice that occur during lapping. These discontinuities enable atoms of materials in the air to enter the lattice and alter the overall properties of the crystal very slightly. For this reason, crystals are enclosed within a sealed container that is either evacuated or filled with an inert gas. The type of can also has an effect. Those with a
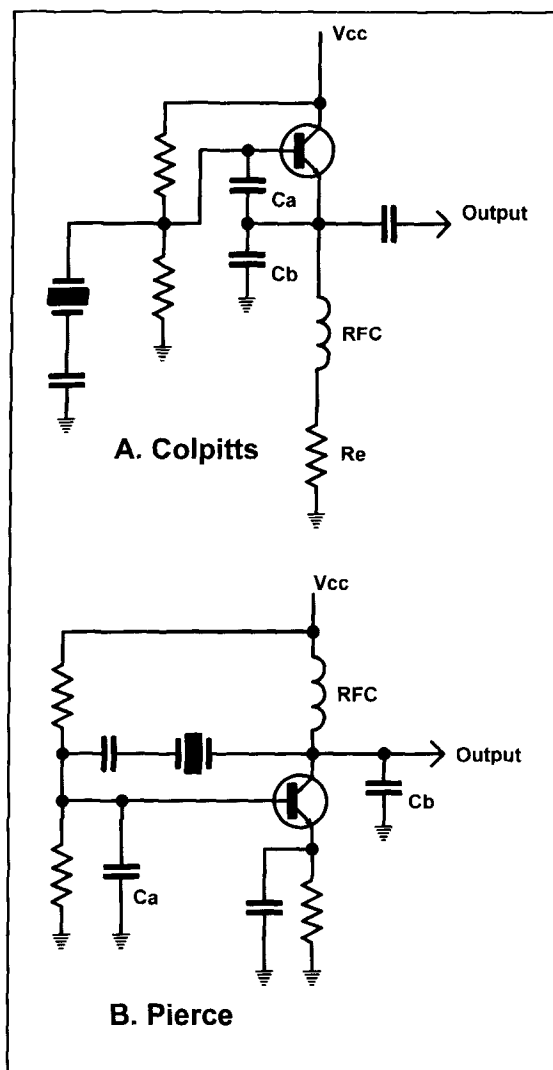
Figure 6. Colpitts and Pierce oscillator circuits.

metal can that has a solder seal are generally the worst, typically giving figures of around 10 ppm per year. Those with a resistance weld generally give figures up to 4 ppm per year, while those with a cold weld on the can give only around 2 ppm per year. Crystals enclosed within a glass envelope provide the best performance with figures of around 1 ppm per year. These figures are included to provide a rough guide to the figures that may be expected.

The way in which the crystal blank is lapped and finished also has an effect. To reduce the size of the discontinuities in the crystal lattice, very fine lapping powder is used. In many cases, in the final stages of manufacture, the crystal is etched. This prevents the abrasive effects of lapping from introducing the discontinuities into which unwanted atoms can enter.

Although these measures reduce the effect of aging quite considerably, some still occurs, and this must be taken into consideration when cal-
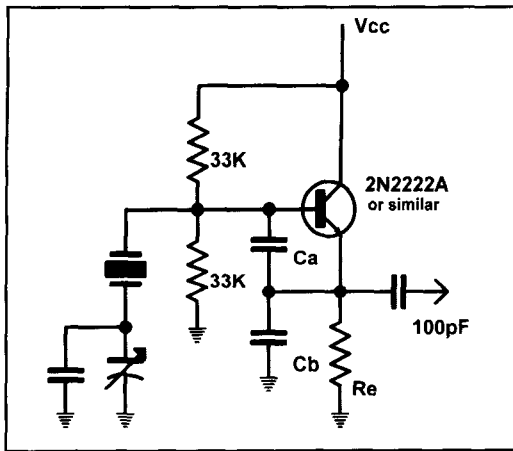
Figure 7. Practical implementation of the Colpitts oscillator.
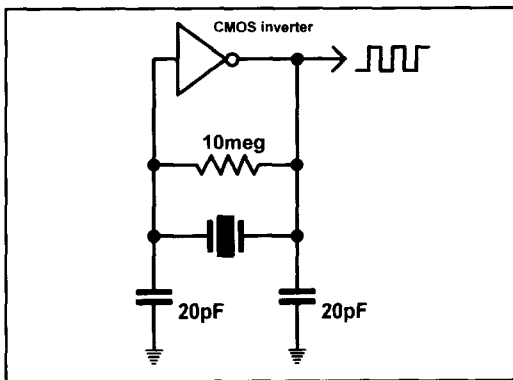
Figure 8. A crystal-controlled oscillator using a CMOS logic gate.

culating the accuracy of the crystal. As might be expected, the degree of aging is much greater in the initial months after manufacture. However there's always some degree of aging, even many years later.

Use in oscillators

Crystals are basically a resonant circuit, and any circuit used must conform to the standard requirements for an oscillator: that the loop gain is greater than unity at the frequency of operation and the phase is an integral multiple of 360 degrees. A wide variety of circuits conform to these requirements. Some are the more traditional types such as the Colpitts and Pierce designs found in radio frequency circuits. However, crystals are widely used as clock oscillators in logic circuits and especially in computers. Here it is more convenient to use logic gates for the circuit.

The most popular discrete circuit is the Colpitts oscillator, although the Pierce circuit is

equally applicable. In both cases, Ca and Cb (**Figure 6**) provide the feedback path, and it can be seen that they are almost identical— except that one is a common emitter circuit while the other is a common collector form. In practice, the Colpitts oscillator is easier to use because one terminal of the crystal can be grounded, making switching between different crystals much easier. However the Pierce circuit has the advantage in that the transistor bias chain only shunts Ca. In the Colpitts circuit, it shunts both Ca and Cb and has a greater effect on the Q of the circuit. Although bipolar transistors do perform well, FETs have a higher input impedance and overcome this problem to a greater degree.

When designing a crystal oscillator circuit, some precautions must be borne in mind. One is to ensure that the correct drive level is maintained. If too much drive is applied, spurious resonances may be excited, giving unwanted signals. In addition, some low-frequency crystals can be damaged if excessive levels of drive are applied, and VHF crystals may be pulled off frequency. In some instances the frequency may be permanently altered, especially if the crystal is operated with excessive drive for extended periods. To ensure that there is a sufficient gain margin, a level of two or three times the level required to maintain oscillation with a crystal at the limit of its activity is normally quite adequate.

Although the basic circuit can be the same whatever crystal is used, some alteration of the circuit values is generally needed to customize the circuit for the frequency of operation. The ratio of Ca to Cb is generally maintained at a ratio of approximately 3:2. Typically, Ca will be about 150 pF at 10 MHz, although it will be lower at high frequencies and higher at lower frequencies. The value of the emitter resistor (Re) shown in **Figure 7** must also be altered. Higher current must be passed through the transistor at higher frequencies and values of around 4.7 k are usually suitable for operation at around 10 MHz, rising slightly for lower frequencies and falling to around 2 k for frequencies up to 25 MHz.

A version of the Pierce oscillator is often used for digital circuits. Here a single CMOS inverter plays the part of the oscillator in conjunction with a minimum of other components. To ensure the circuit operates correctly, it's wise to follow the oscillator with a buffer—otherwise the inputs of the gates being driven or long tracks may cause an undue level of loading.

Although the performance of a crystal may appear to be overkill for a clock oscillator for a digital circuit, a crystal can provide a very cost-effective and reliable solution. Often crystals for clocks can be bought in quantity for several

tens of cents, and this makes them a very attractive proposition.

Filters

Crystals are widely used in high-performance filters for equipment like radio receivers and transmitters. Here, their exceedingly high Q values mean that the performance of crystal filters can easily exceed other versions, especially the more common LC types.

It is not easy to produce a filter with the required shape with a single crystal. For most applications a response curve like that shown in **Figure 9** is required. This enables a certain bandwidth to be passed without attenuation of some of the wanted frequencies. A single crystal won't exhibit a characteristic of this nature. Instead it will have a much narrower pass-band, and its ultimate rejection (i.e., the final rejection away from the pass band) may not be sufficient. The other problem is that the response isn't symmetrical, and again this is not ideal. To achieve the required bandwidth and overall performance, a number of crystals are used together in a configuration like that shown in **Figure 10**.

The filter in **Figure 10** is symmetrical and provides a wider bandwidth, often a few kilohertz, that is ideal for receiving transmissions off the air. Amplitude modulated (AM) signals usually require a bandwidth of around 6 kHz (for short wave transmissions), single sideband (SSB) around 2.7 kHz, and Morse 500 Hz or less, depending upon the levels of interference.

To achieve a wider bandwidth than that obtained with a single crystal, the two crystals used in the configuration should have slightly different frequencies. The 3-dB bandwidth of the filter (i.e., the bandwidth where the response of the filter has fallen by 3 dB) is about 1-1/2 the difference between the resonant frequencies of the two crystals. Also, the response has a peak at either side, with a dip in the middle. To reduce the dip, the series resonant frequency of one crystal should be the same as the parallel resonance of the other.

The filter shown only has two poles (i.e., resulting from only two crystals). This won't normally provide the performance required for modern communications receivers, and further stages are generally added to give the required performance. Generally six or eight poles are common for high-performance communications receivers. I won't expand upon this here because most readers will be very familiar with filters found in receivers and transceivers.

Filters have a certain input and output impedance. For a crystal filter to operate correctly, it must be driven from a source with the correct impedance and matching into a load with the
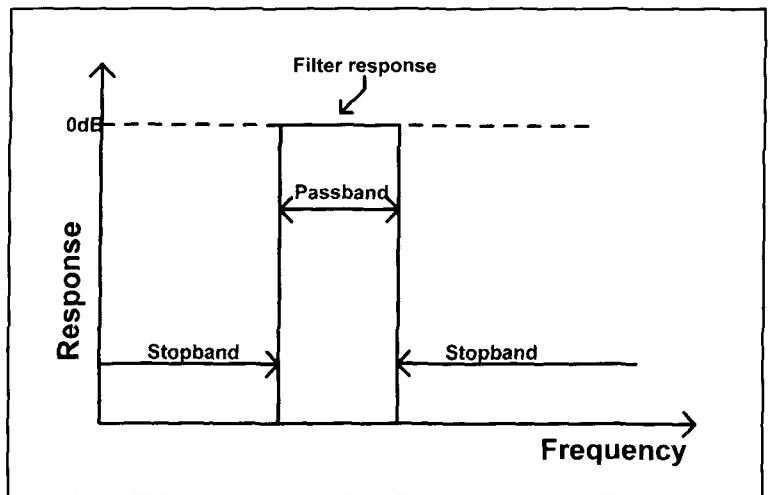
Figure 9. An ideal filter response.
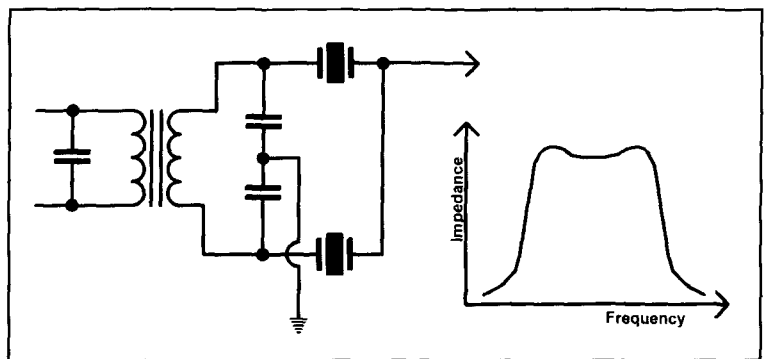
Figure 10. A half lattice crystal filter configuration.

correct impedance. If there's a mismatch, the response will be degraded, as will other parameters like the ripple. To ensure the correct match is present, resistors are often placed across the input and output.

Monolithic filters

Just as the trend in the semiconductor industry has been to place as many components onto a slice of silicon, the same is true of crystals—but to a much lesser degree. Instead of requiring that several individual crystals be contained in a can, a monolithic filter uses a single quartz crystal.

The way in which a monolithic filter works wasn't fully understood for many years. Even though it had been realized that a filter could be made on a single quartz crystal, the idea wasn't developed. It wasn't until research work at Bell Laboratories in the U.S. was undertaken that a sufficient understanding was gained to manufacture these filters.

A monolithic filter is normally fabricated on an AT cut crystal to provide the optimum per-
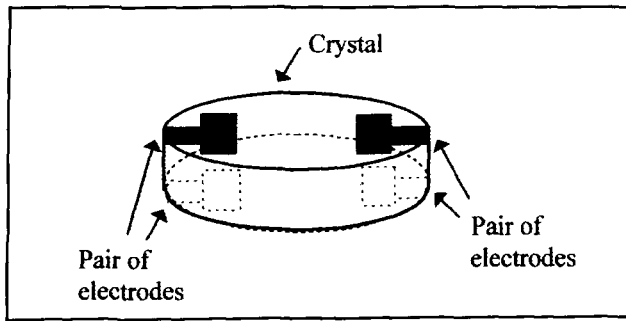
Figure 11. A monolithic crystal filter.

formance in terms of activity, frequency coverage capability, and temperature stability. It's made by placing pairs of electrodes at opposite ends of the crystal as shown in **Figure 11**. The coupling between these electrodes acts in such a way to provide a highly selective filter.

The filter operates in a thickness shear mode. The resonant frequency is proportional to the thickness of the piezo material such that the thickness of the material is $n\lambda/2$, where λ is the operating wavelength and n is an odd integer. The filter acts as two acoustically coupled resonators, thereby giving an improvement in performance over a single crystal filter.

One set of electrodes acts as the input and the other as the output. Signals at the input are converted to mechanical vibrations that are affected by the mechanical resonances in the crystal. The vibrations travel across the crystal affected by the resonances in the resonant regions shown. Obviously those signals that do not match the resonance of the filter are severely attenuated. Finally, the signals within the passband of the filter are converted back to electrical signals at the output electrodes.

The electrical properties of a monolithic filter can be described in terms of an equivalent circuit, as shown in **Figure 12**. In this circuit the two sets of elements La1, Ca1 and La2, Ca2

determine the resonant frequency of the filter. The dimensions of the crystal determine these, and as a result they need to be controlled very accurately to ensure the correct performance is achieved. Also, the two sections are chosen to have their resonant frequencies very slightly different. This ensures that the correct bandwidth is achieved in the same way that it was for discrete crystal filters.

A number of other elements are also present. Lb is the internal coupling between the two resonators; Cb forms the input or output capacitance, and it equates to the capacitance between the electrodes at one end. One of the most crucial elements is Cp—the parasitic capacitance between the two sets of electrodes. Signals taking this path bypass the filter, therefore the value of this element must be kept to an absolute minimum. High values of Cp will mean that out-of-band signals won't be sufficiently rejected, and the performance may not be satisfactory.

Monolithic filters are generally manufactured for frequencies below about 40 MHz. Above this, the crystals become very thin and the yield in manufacture reduces. However if higher frequencies are required, the filters can be designed for operation in an overtone mode. In this way the maximum frequencies of operation can be increased considerably, although there is a small penalty in performance degradation.

Most filters used within communications filters have either six or eight poles. To gain greater degrees of selectivity, it's possible to add further electrode pairs to the same substrate. By using the same substrate for the filter, manufacturing costs can be reduced and the performance increased.

Despite the advantages of monolithic crystal filters, relatively few companies manufacture them. Specialized techniques are required in their manufacture and testing, and this entails a considerable investment. Additionally, the required quantities of monolithic filters are relatively low compared with the crystals that are needed in many very high volume products including cellular phones, computers, televisions, and so forth.

Summary

While crystal technology may not appear to be as up to date as semiconductor manufacture, this is not the case. Crystals are being used increasingly as very high-performance, low-cost elements in a very wide variety of circuits. Development is taking place in a number of areas, and as there are no obvious contenders to take over their market supremacy, they will be widely used in radio and electronics for very many years to come. ■
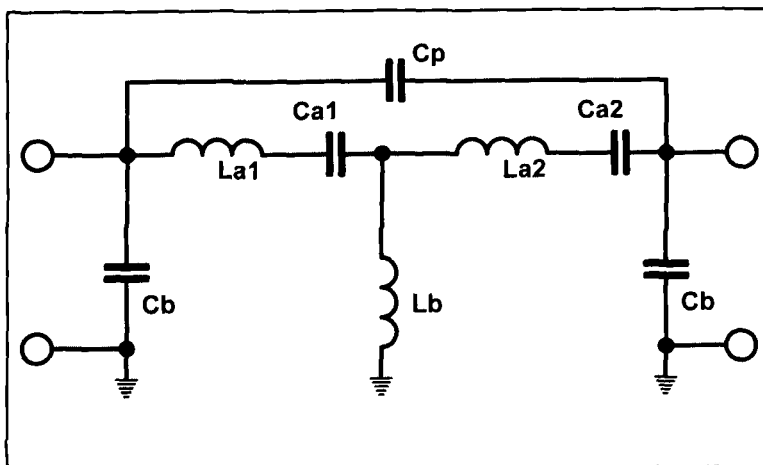
Figure 12. Equivalent circuit of a monolithic crystal filter.

R.P. Haviland, *W4MB*
1035 Green Acres Circle, North
Daytona Beach, Florida 32119

THE QUAD ANTENNA REVISITED

Part 1: Developments in antenna analysis

It has been nearly 20 years since I started to study the quad antenna family in detail, and just over 10 years since the first publication of the study results.[1] During this time there have been developments in the field of antenna analysis that provide much improved data, cal-culated values which come much closer to values found in test. There has also been a great increase in computer capacity and speed: evaluations that took 10 or 12 hours at the start of the study now take seconds. In addition, some techniques of transfer from "antennas on
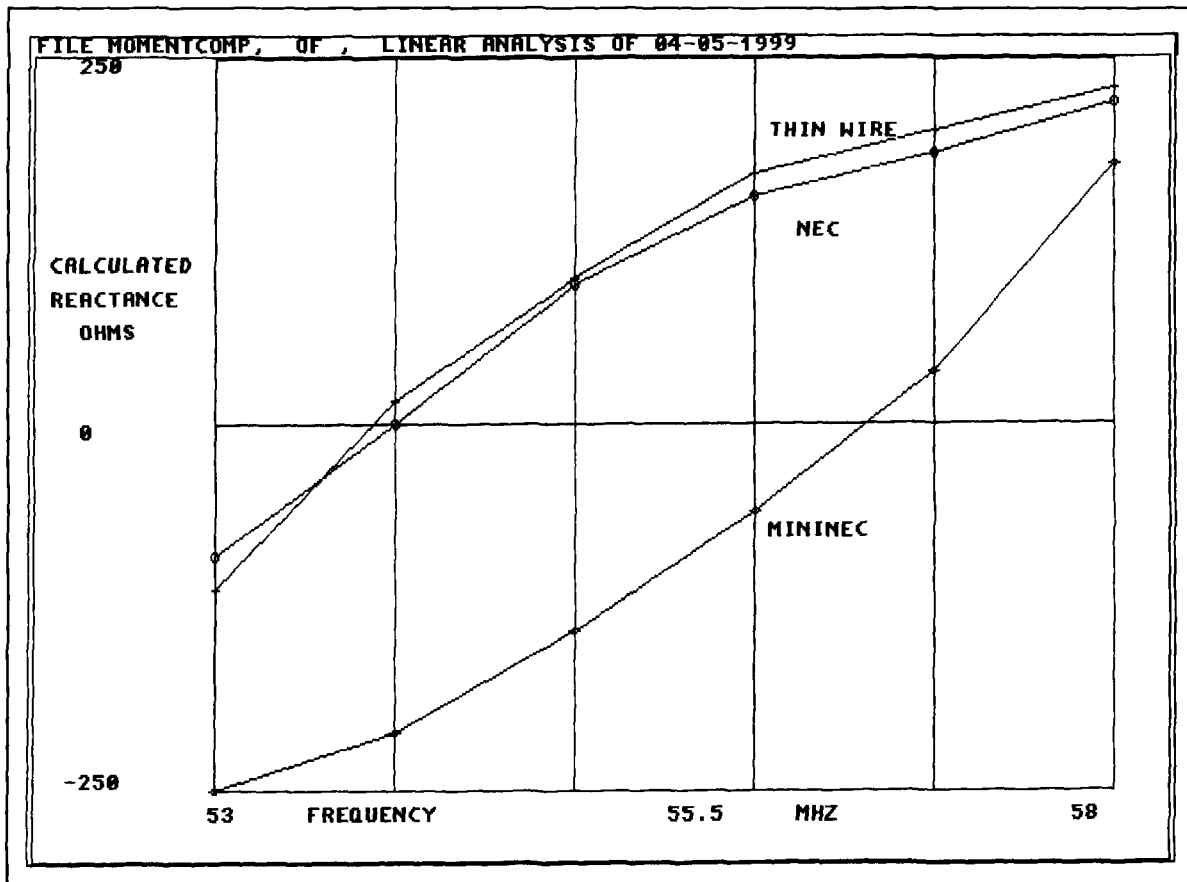
Figure 1. Comparison: MININEC, NEC, and Thin Wire analysis.
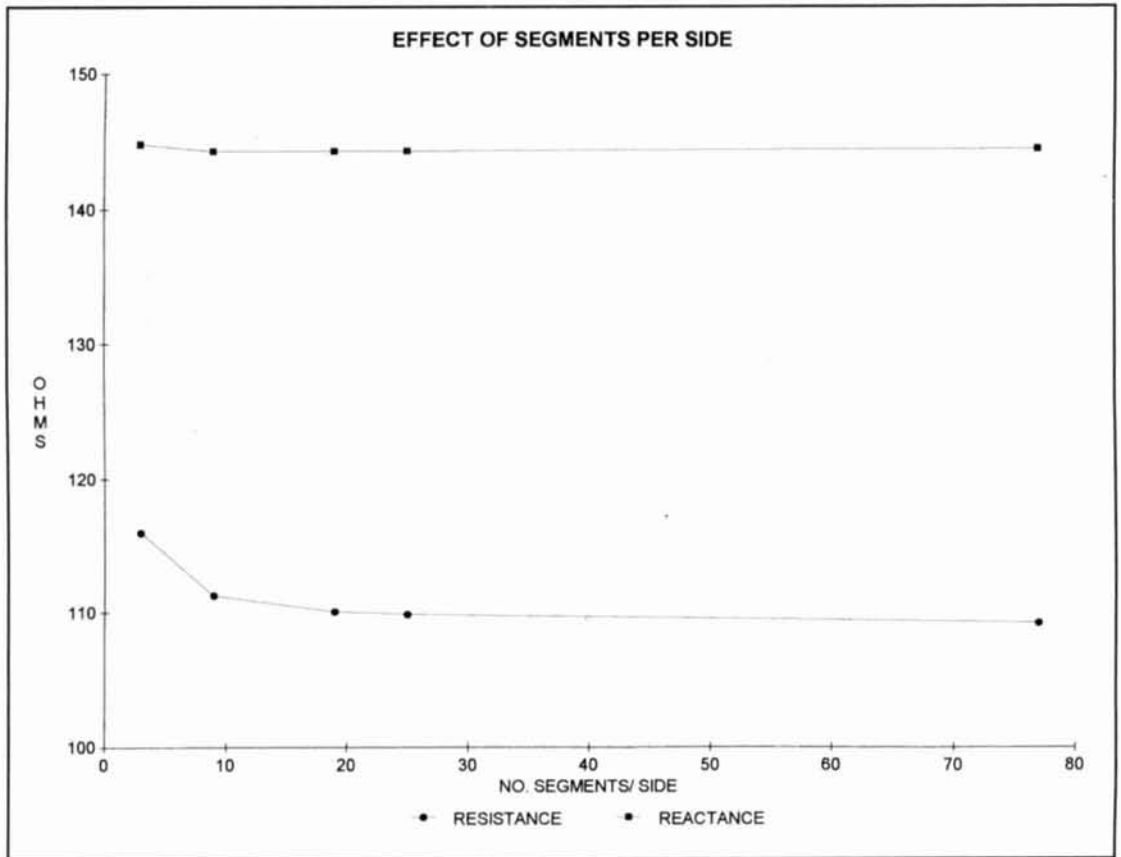
Figure 2. Effect of wire segmentation on impedance.
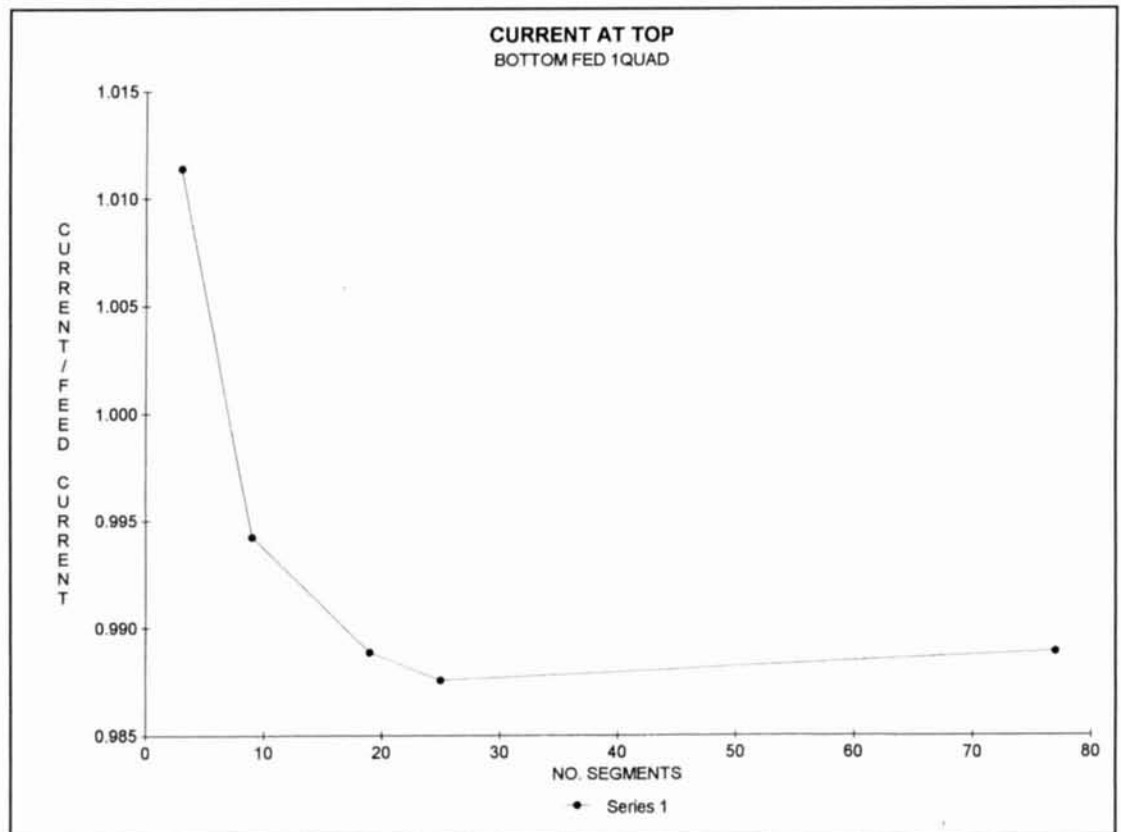
Figure 3. Effect of wire segmentation on current.
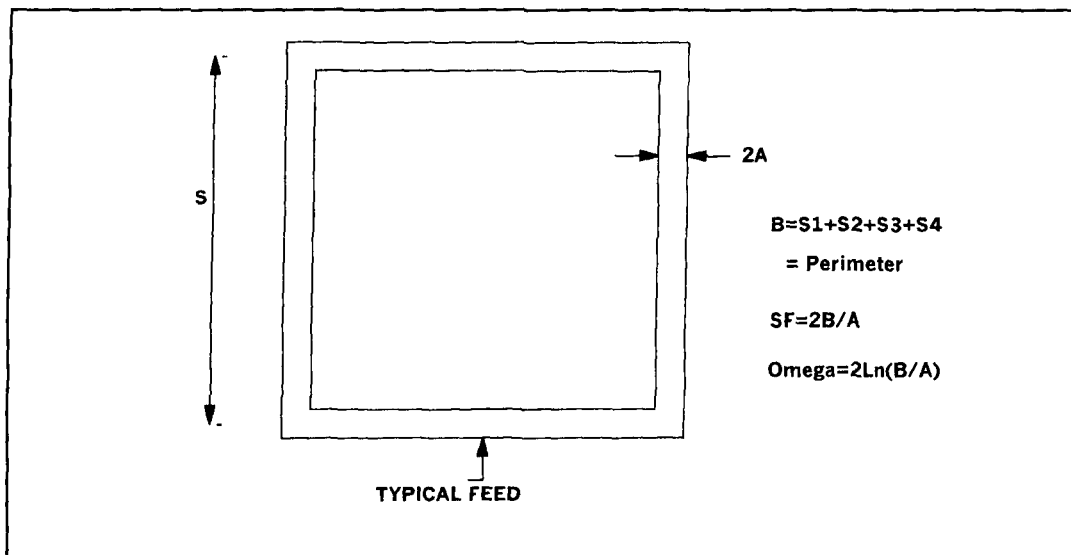
Figure 4. The basic quad loop.

paper" to hardware antennas in the real world have appeared, which make it easier to obtain the design performance.

These factors have added up to the point that it seems desirable to revise the results of my studies. published in the book *The Quad*

Antenna.[2] The overall goal is to improve accuracy, to provide data not possible with earlier analysis methods, and to make the results easier to use.

This series of articles will concentrate on the changes that have developed. The plan is to
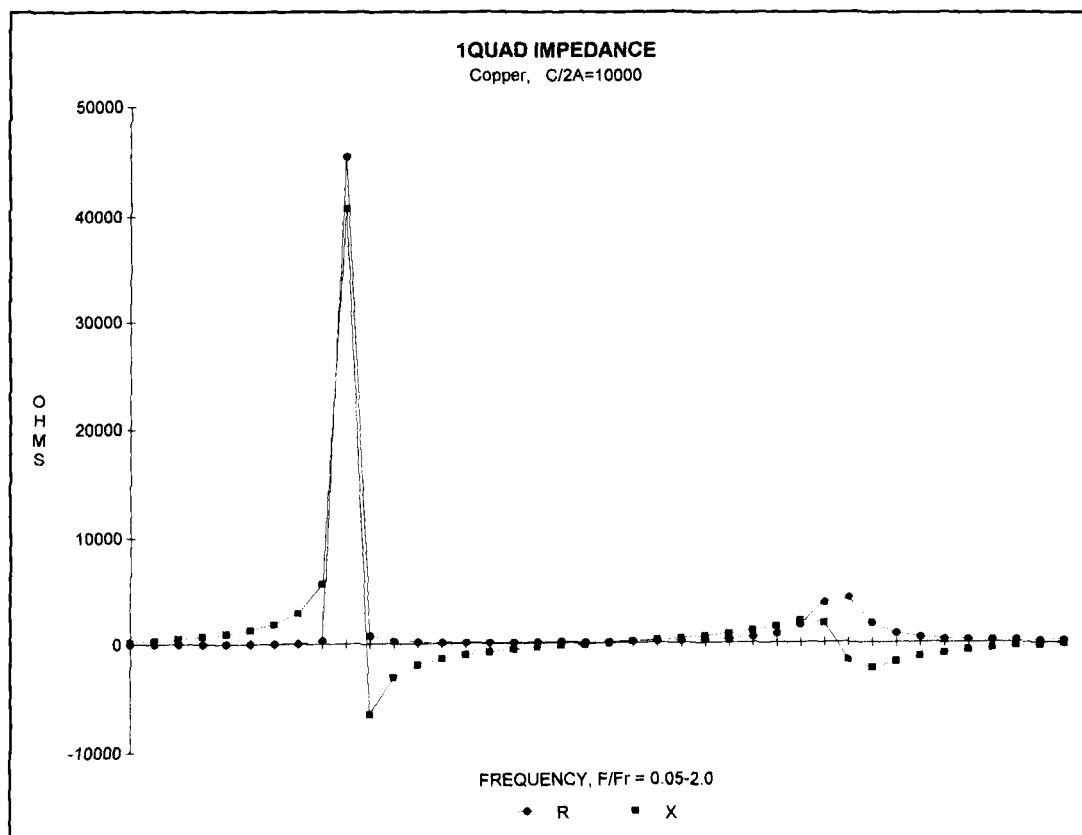
Figure 5. Quad loop impedance versus size, thin wire.
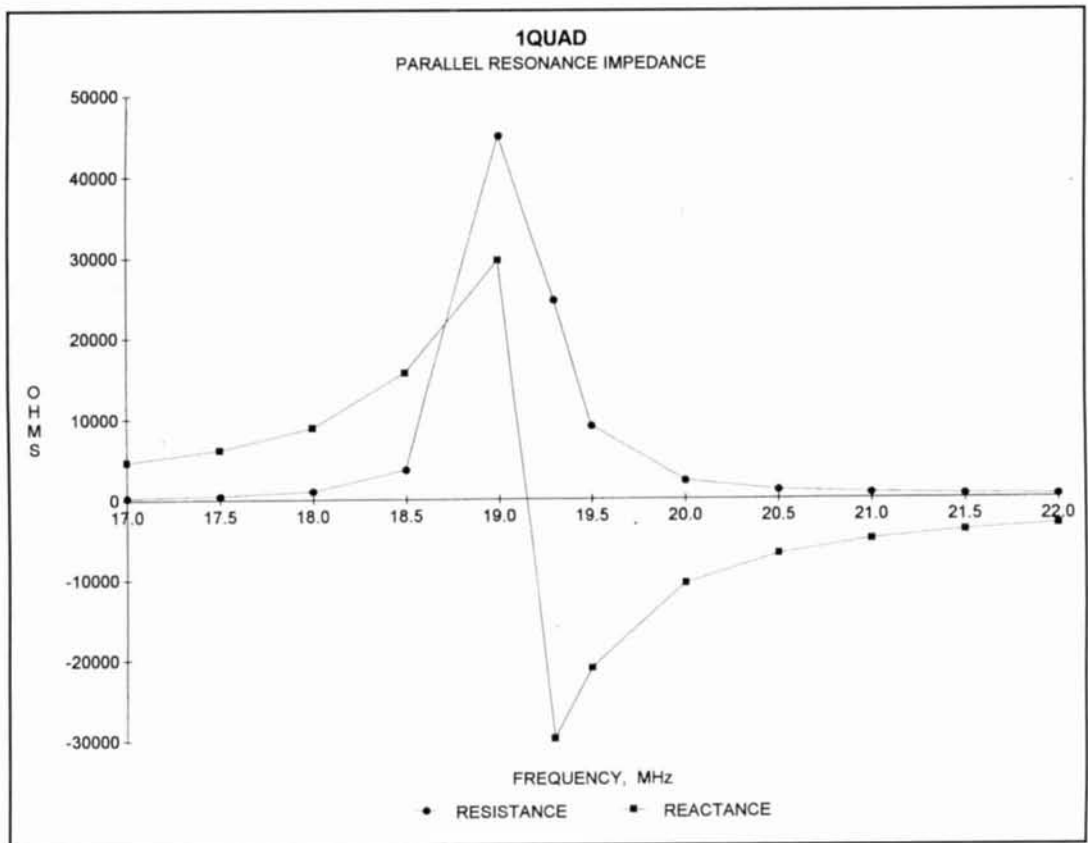
Figure 6. Quad impedance near parallel resonance.
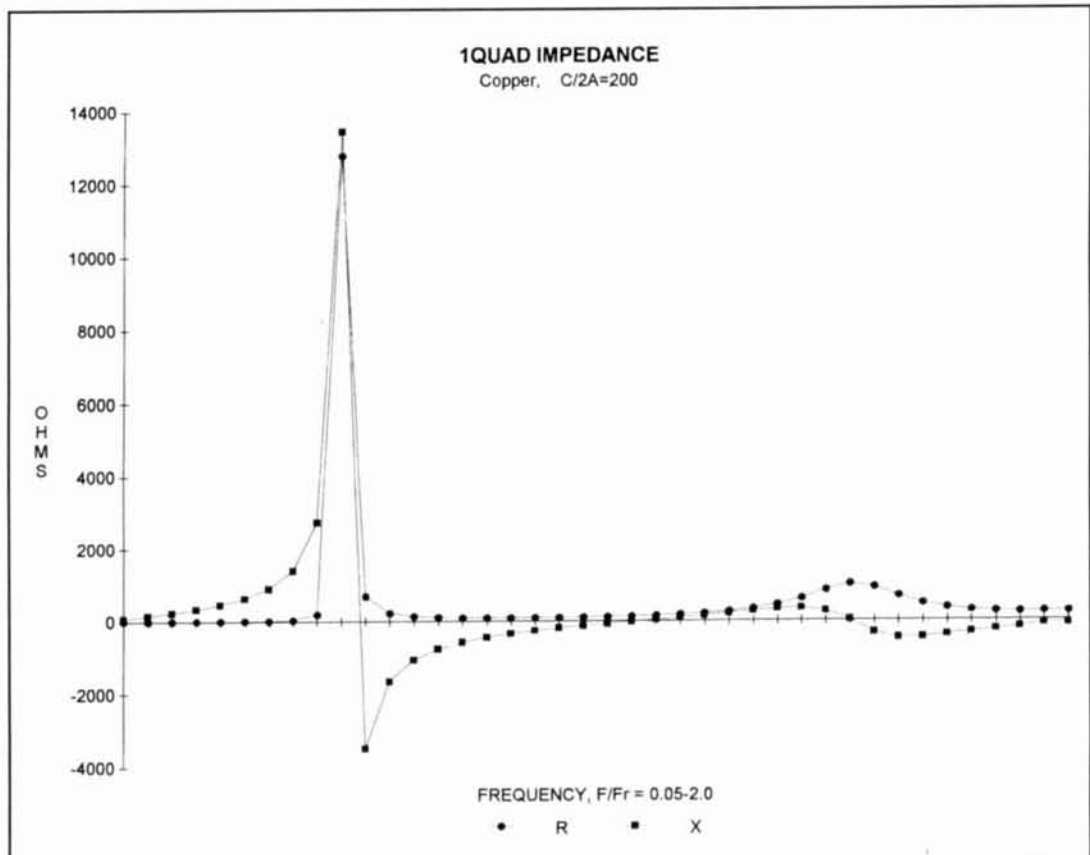
Figure 7. Quad loop impedance versus size, thick wire.

Table 1. Quad loop impedance versus loop size.

kB	freq	100 R	X	200 R	X	500 R	X	1000 R	X	3000 R	X	10000 R	X	30000 R	X
0.05	1.875	0.01	60	0.02	73	0.06	90	0.12	104	0.35	125	1.2	148	4.4	173
0.1	3.75	0.03	123	0.05	150	0.1	185	0.18	212	0.53	256	1.8	303	5.9	350
0.15	5.625	0.1	195	0.12	235	0.19	290	0.31	333	0.77	400	2.5	476	7.8	547
0.2	7.5	0.35	282	0.36	335	0.5	414	0.61	474	1.2	569	3.6	703	10.4	780
0.25	9.375	1.1	395	1	462	1.2	569	1.4	650	2.2	781	5.3	934	14.6	1073
0.3	11.25	3.4	557	3.1	635	3.3	780	3.6	891	4.8	1069	19.3	1285	22.8	1475
0.35	13.125	11	829	9.4	903	9.7	1103	10	1260	12	1508	20	1828	41.9	2180
0.4	15	49	1410	33	1400	34	1696	35	1930	39	2300	57	2846	102	3275
0.45	16.875	508	3701	190	2733	183	3245	184	3671	197	4300	291	5673	456	-6554
0.5	18.75	2169	-6011	12790	13450	9830	17050	9485	19600	10030	24000	45520	40740	58640	-34670
0.55	20.625	255	-1797	687	-3489	801	-4565	850	-5312	899	-6401	740	-6528	891	-7291
0.6	22.5	125	-1045	223	-1658	244	-2090	253	-2398	263	-2865	244	-3127	283	-3522
0.65	24.375	89	-728	136	-1066	146	-1325	150	-1512	155	-1797	148	-2002	166	-2257
0.7	26.25	75	-549	105	-767	112	-943	115	-1071	118	-1267	114	-1423	126	-1602
0.75	28.125	69	-431	92	-581	97	-703	100	-799	102	-939	100	-1059	108	1188
0.8	30	68	-345	87	-451	92	-542	93	-607	96	-709	95	-798	101	-891
0.85	31.875	70	-278	86	-351	91	-415	92	-461	95	-532	94	-596	100	-660
0.9	33.75	74	-222	89	-271	93	-312	95	-341	97	-386	99.1	-424	103	-468
0.95	35.625	80	-174	94	-201	98	-222	100	-237	106.5	-259	103	-279	109	-298
1	37.5	88	-132	102	-140	106	-142	108	-143	110.3	-149.2	113	-142	119	-138
1.05	39.375	100	-92	112	-82	117	-67	120	-53	122.7	-32.9	126	-9.5	133	16
1.1	41.25	115	-55	126	-27	132	7.1	135	34	139.1	76.8	145	125	153	173
1.15	43.125	135	-19	145	27	153	82	157	124	162	191	171	266	181	340
1.2	45	161	16	170	82	181	160	186	219	193	314	208	425	221	526
1.25	46.875	197	49	205	140	221	245	229	326	240	453	265	607	283	747
1.3	48.75	245	78	254	200	280	341	294	449	312	621	356	835	385	1023
1.35	50.625	310	96	328	282	373	450	399	595	434	829	520	1135	570	1397
1.4	52.5	395	89	440	318	529	569	586	770	661	1100	857	1552	966	1936
1.45	54.375	489	36	609	340	806	666	746	945	1152	1428	1688	2087	2014	2685
1.5	56.25	551	-77	835	256	1285	590	1630	904	2257	1502	3837	1853	5036	2368
1.55	58.125	531	-215	1001	15	1688	27	2290	14	3466	-136	4314	-1649	5063	-2614
1.6	60	440	-304	914	-346	1391	-668	1729	-1016	2158	-1689	1852	-2389	1917	-2896
1.65	61.875	339	-326	669	-492	865	-825	959	-1097	1040	-1519	862	-1785	877	-2067
1.7	63.75	262	-306	466	-477	541	-715	569	-889	587	-1146	503	-1305	513	-1492
1.75	65.625	213	-270	343	-407	377	-570	388	-685	394	-855	350	-971	359	-1101
1.8	67.5	185	-232	274	-331	293	-442	300	-521	304	-637	279	-724	287	-816
1.85	69.375	172	-196	239	-261	252	-336	257	-388	260	-467	245	-530	253	-593
1.9	71.25	170	-163	223	-199	234	-246	238	-278	242	-327	233	-369	241	-401
1.95	73.125	175	-135	229	-94	231	-167	235	-183	239	-306	235	-228	243	-241
2	75	187	-112	228	-94	239	-98	244	-98	248	-97	249	-98	258	-90
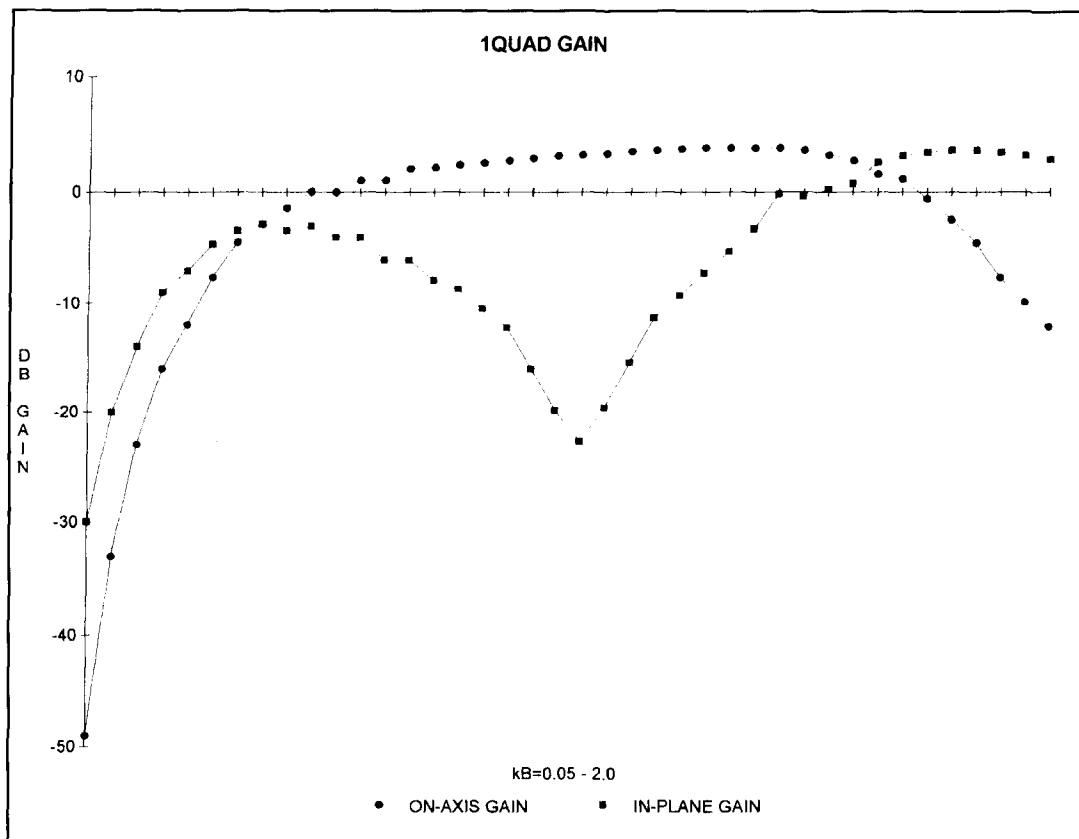
Figure 8. Gain of quad loop, on-axis and in-plane.

integrate these with the published book, as a "second edition, revised and corrected."

Developments in antenna analysis

The fact that all antenna analysis is based on equation approximations must be accepted. Even many of these must involve further approximations to get numerical values.

The original quad studies were made with the public domain program "MININEC",[3] followed four years later by MININEC3. Both used two simple approximations: the antenna is assumed to be formed of connected short segments of wire, with the current constant over a segment. To account for the fact that current must be zero at the end of the wire, half segments of zero current are assumed at the end. The assumptions work quite well for such antennas as dipoles and Yagis. Unfortunately, they are less reliable for antennas that have zero currents at point(s) away from the ends, and even less valid for antennas with no end—the whole family of quads. (There is a new commercial version of MININEC available, with some changed assumptions to eliminate

these problems. The new program is also easier to use.)

Once there existed a program with different assumptions, but it was restricted to military use. A major difference was that the current on a segment was represented as the sum of a constant plus a sine term plus a cosine term. This greatly improved accuracy. The program later became available, and is currently known as NEC2.[4] It is also the basis of several commercial versions, simplified for easier use, but having full NEC accuracy for most antennas (see ads in amateur magazines). One of these commercial versions is used for many calculations here. Further improvements and extensions are available in NEC4, now available to U.S. citizens, but not inexpensively. This has not been used in the extended study.

Still another analysis program is "Thin Wire".[5] This assumes that the currents are sinusoidal sections, which can simulate triangles at the end of a wire, or essentially constant current, or most changes encountered in antennas.

MININEC, NEC, Thin Wire, and others are available in the commercial CD-ROM, *Antennas*. It is almost always available at hamfests.

The general result of the various assumptions is that all programs agree quite well with measurements far from the antenna, such as the far-
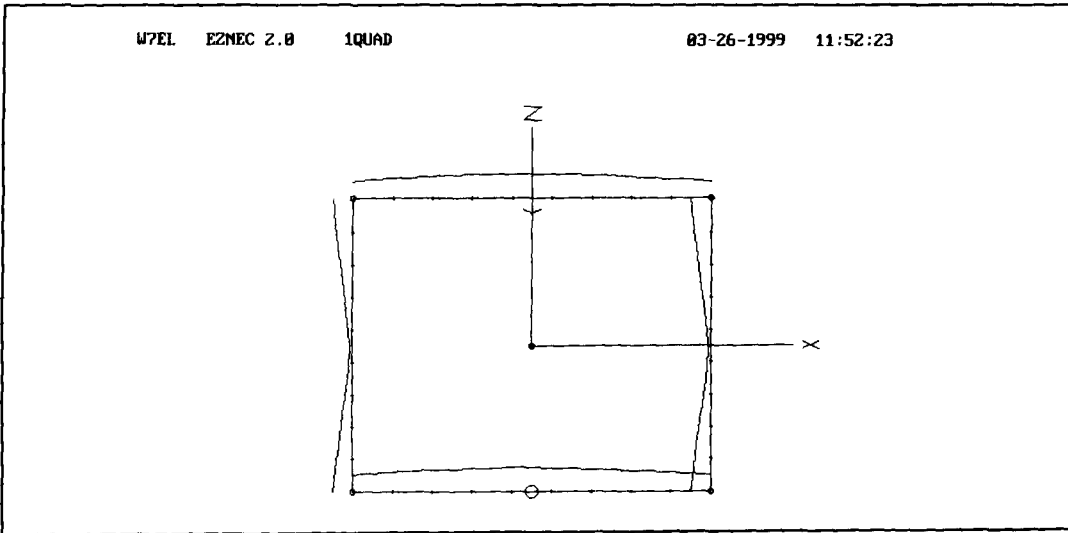
Figure 9. Current on loop near series resonance.

field pattern. Differences become larger for variables close to the antenna, such as drive resistance and reactance. This is especially true for MININEC. (See the comparison of calculations for a typical small quad shown in **Figure 1**.)

A result is that the performance calculations in the original quad book are valid for most quantities; but some critical ones, such as best F/B ratio, occur at a different frequency than

shown from the MININEC calculations. Correction of these is a goal of this revision.

Effects of newer computers

As a consequence of memory and speed limitations, the quad book calculations were based on four segments per side, usually 16 per wave-
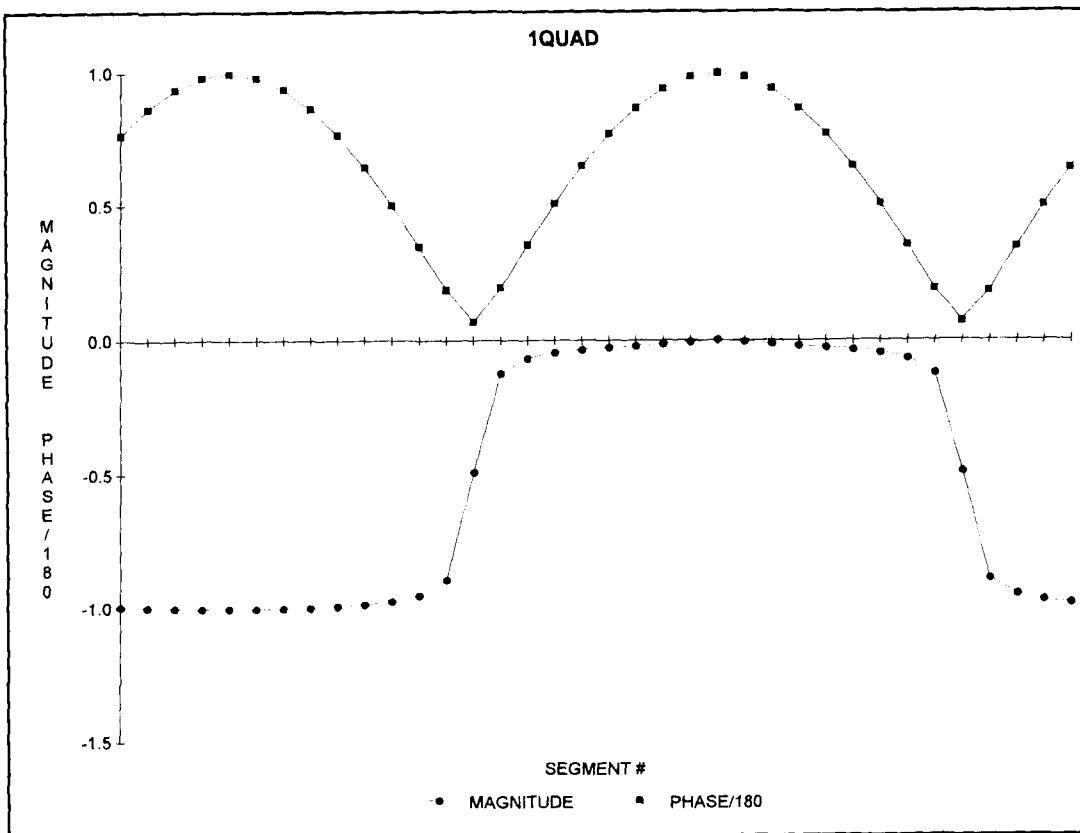
Figure 10. Current and phase around loop near resonance.

1QUAD   EZNEC 2.0

03-26-1999 14:26:48
Freq = 40 MHz

Tot————
H ....
V - - - - - - -

Cursor = -21.46 dBi
       = -24.76 dBmax
Outer Ring = 3.30 dBi
Max. Gain = 3.30 dBi

                    0 deg.

Gain: 3.30 dBi
Angle: 90 deg
F/Side: 24.76 dB
Bmwidth: 84.2 deg
-3: 47.9, 132.1 deg
Slobe: 3.30 dBi
Angle: 270 deg
F/Slobe: 0.00 dB

Cursor Az =   0.0
Azimuth Plot
Elevation Angle =   0.0 deg.

1QUAD   0 dB   EZNEC 2.0

03-28-1999 15:16:55
Freq = 40 MHz

Tot————
H ....
V - - - - - - -

Cursor =   3.30 dBi
       =   0.00 dBmax
Outer Ring = 3.30 dBi
Max. Gain = 3.30 dBi

                    0 deg.

Cursor El =   0.0
Elevation Plot
Azimuth Angle = 90.0 deg.

Figure 11. Quad radiation pattern, near resonance.
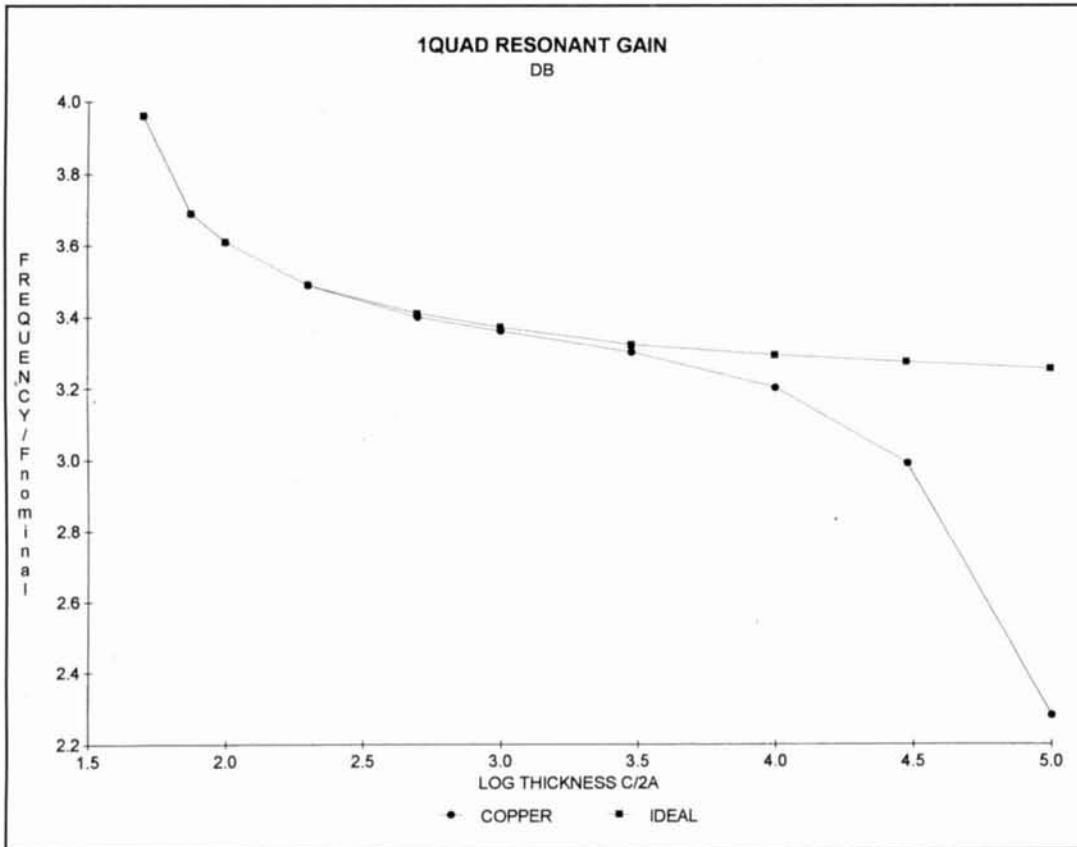
Figure 12. Effect of wire thickness on gain.
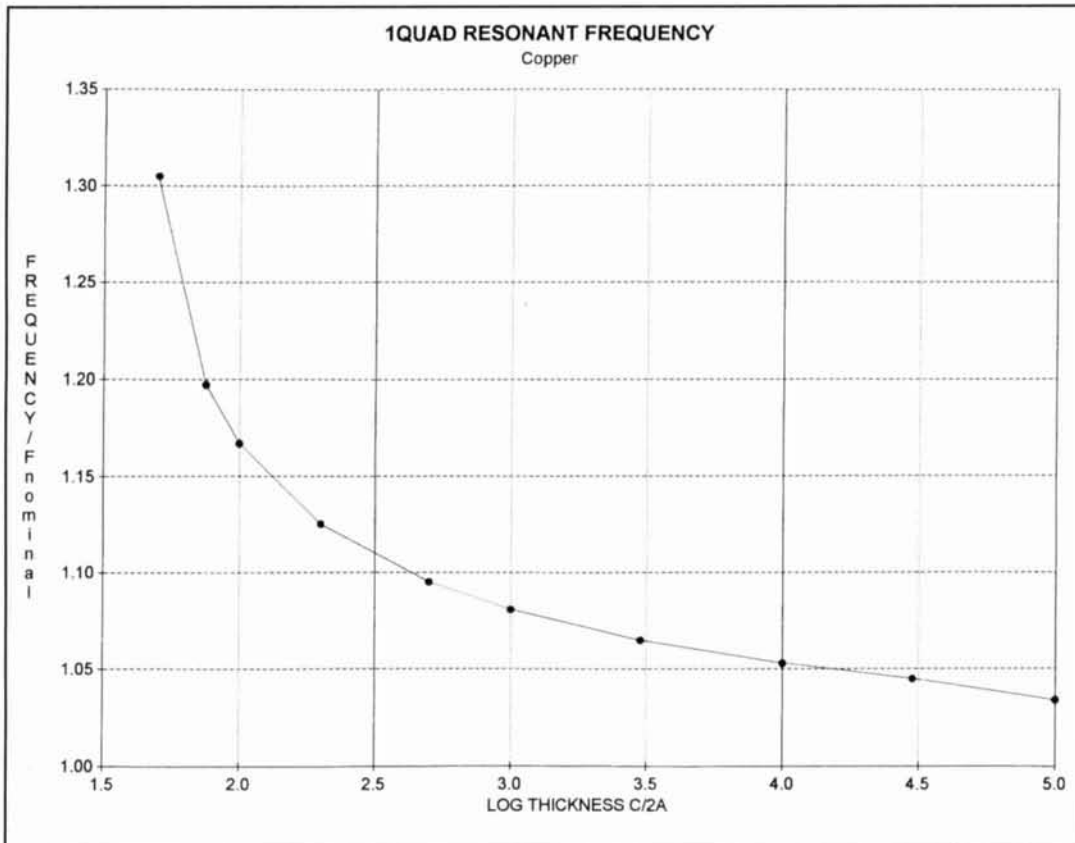
Figure 13. Effect of wire thickness on resonant frequency.
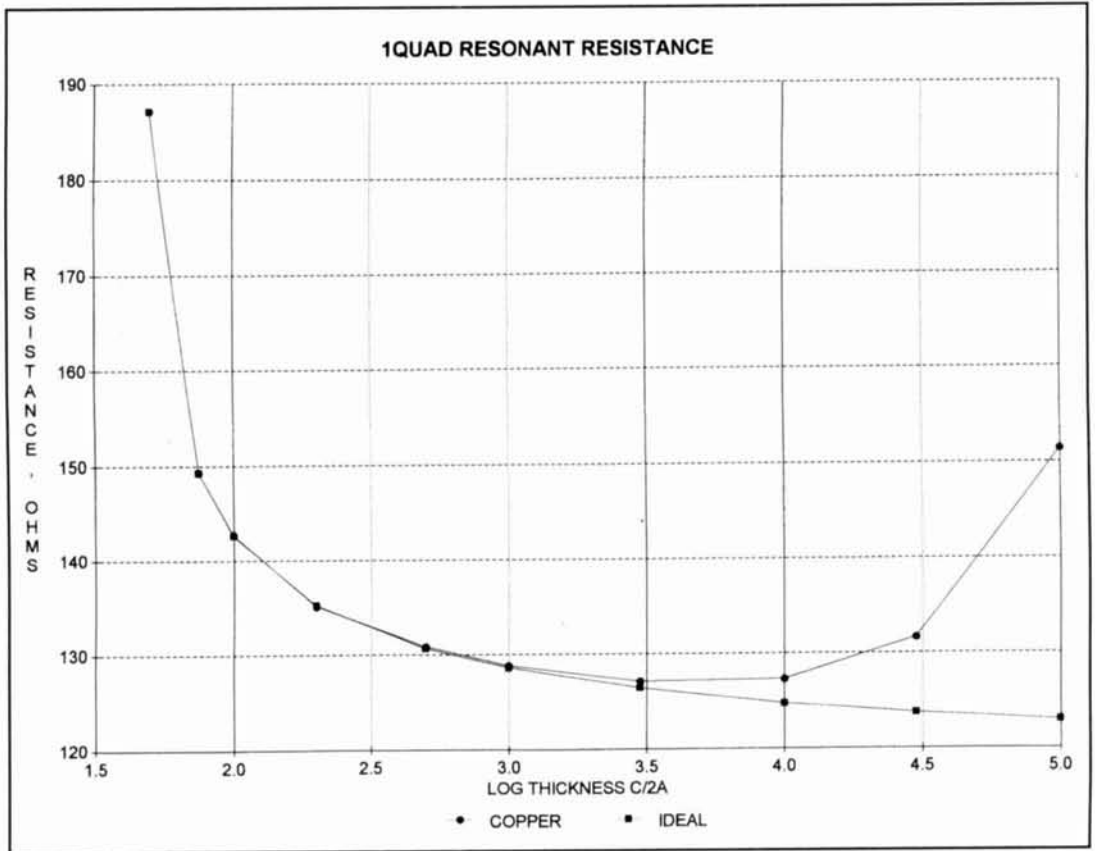
Figure 14. Effect of wire thickness on resistance.
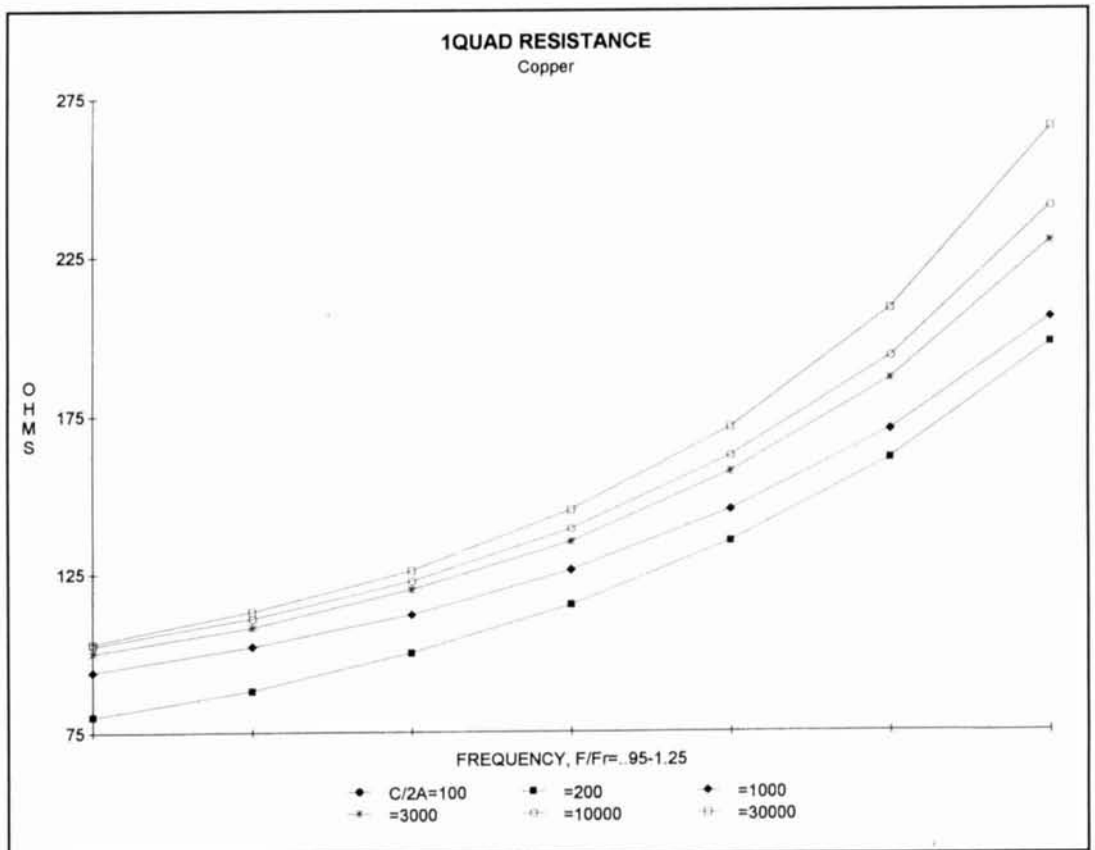
Figure 15. Resistance change with frequency, thickness changing.

1QUAD REACTANCE
Copper

FREQUENCY, F/Fr = 0.95-1.25

- ● P/2A=100
- ■ =200
- ◆ =500
- ✳ =1000
- ⊕ =3000
- ⊟ =10000
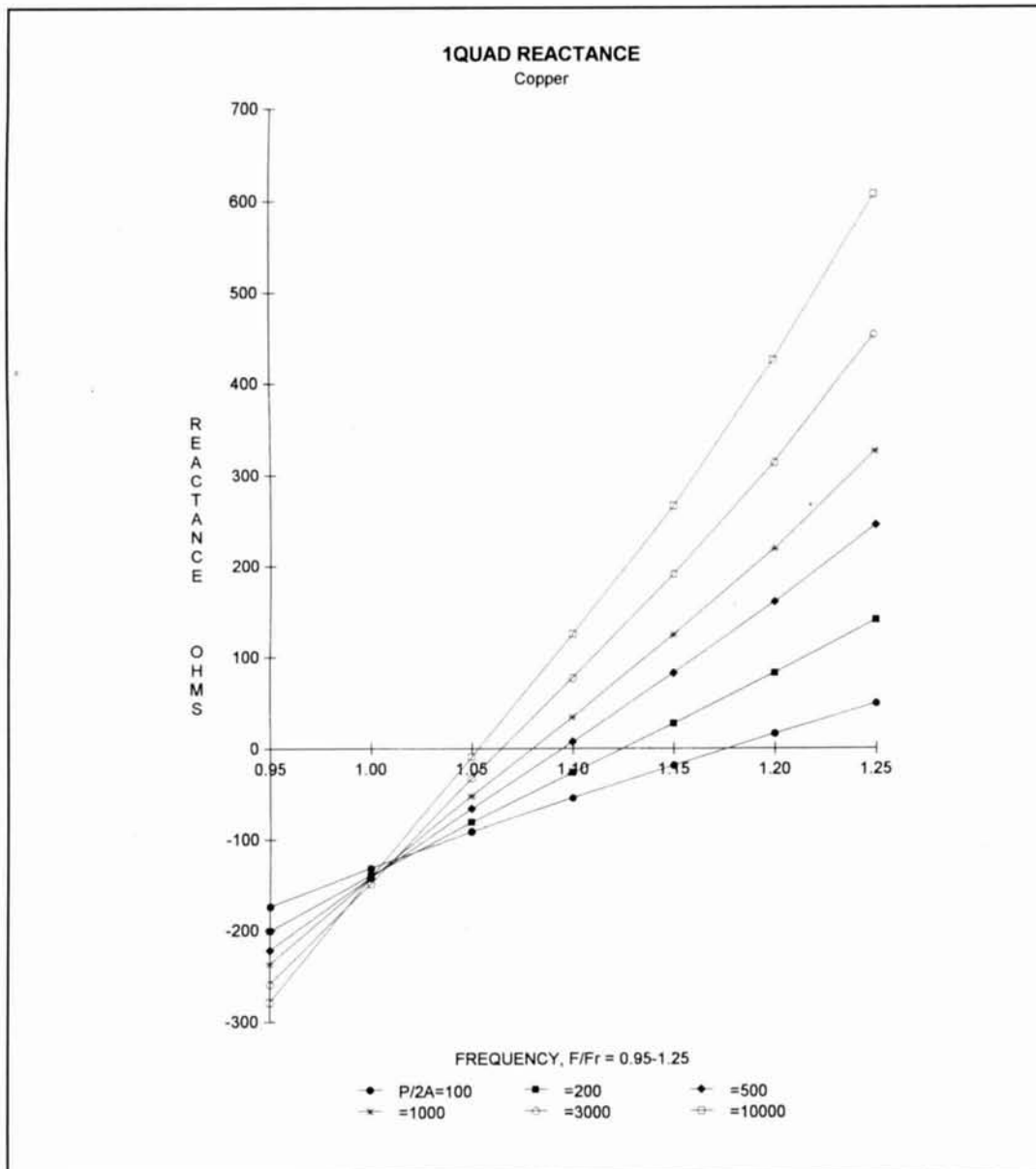
Figure 16. Reactance change with frequency, thickness changing.

length. The usual test for segmentation is to try calculations with increased number of segments: a small or no change is an indication that segmentation is adequate. **Figure 2** shows such a result with NEC. However, note another effect: If the segments become too short, approaching the diameter of the wire, other assumptions become less valid. This appears here as an increase in values with a very large number of segments. Virtually all work in these revisions is with nine segments per quad side, providing reasonable accuracy with NEC.

Figure 3 shows another consequence of insufficient segmentation. With just a few segments, the current at the top of the loop exceeds that at the feed point—a physical impossibility. At 24 segments, the calculated current has

become essentially constant. However, the effect of too small segments also shows.

Increased program capability

MININEC modeled ground as a perfect mirror, or by a "reflection coefficient" approximation, essentially an assumption that the ground reflected signal was equal to the direct signal times a loss factor. These were adequate for most pattern calculations; however, the ground directly under an antenna is always assumed to be perfect. The result was that antenna impedances and patterns were poorly modeled for antennas lower than about 0.2 wavelengths.
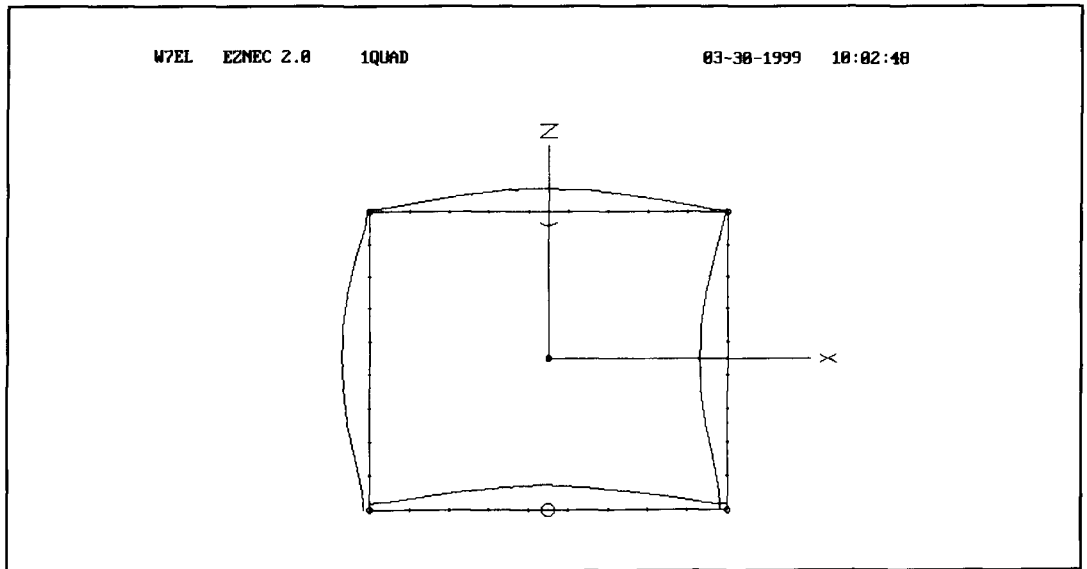
Figure 17. Current on loop at second resonance.

NEC provides perfect and reflection coefficient ground models, as well as an added one based on the exact equations of Sommerfield. While this method starts as exact relations, approximations must be used in the calculations. Also, ground conditions are rarely known to great accuracy. The recommendation is to use the best values available, plus a check of ground importance by varying assumed ground values. If the ground effect is large, a program of ground measurements is indicated.[6] NEC2 allows choice of the ground model under the antenna, so it can give much better results for low antennas. NEC4 allows antennas to be buried.

The original thin wire program was for free space only, but an ideal ground could be modeled by providing an "image antenna" set at twice the antenna height above ground. Adler and students at the Navy postgraduate school added a reflection coefficient method. The program is included in the *Antennas* CD-ROM mentioned above.

MININEC and NEC2 handle bare wire only. The thin wire program and NEC4 allow insulation on the wire. For the work here, a family of approximations is provided to develop a correction factor for insulation.[7] For critical applications one of the two program methods should be used.

General changes

Over time, several small changes have been developed to simplify antenna analysis and analysis application. Probably the most important of these is the realization that it is easy to measure the resonant frequency of an antenna element in the real world, using a grid-dip meter plus a frequency counter for accuracy.[8] Its importance is that this measures the effect of tapered elements, clamping hardware, the boom, and even surrounding objects. Adjusting element size to bring this measured value to that which is calculated means that calculated antenna performance is easier to attain. About the only thing to watch is the effect of antenna height during measurement. If this is greater than 0.2 wavelength, the error is small. If less, the resonant frequency should be calculated for the element to be tested at the test height.

As a consequence of this change, plot and table coordinates used here may differ from those used in the original study. The changes will be reflected in the text.

The basic quad loop

The basic square quad element is sketched in **Figure 4**. This defines the dimensional quantities used in analysis. The effect of wire size can be shown in two ways. The factor "omega" is much used in theoretical studies, because of the form of some equations. However, for practical use, the ratio of wire length to diameter is simpler. In this nomenclature, its value is P/2A. At HF, a value in the thousands indicates a fine wire, and one in the hundreds a tube or its equivalent, a cage of wires.

The quantity kb can be interpreted in two ways. If b is assumed constant, it measures the exciting signal wavelength. If wavelength is assumed constant, kb measures the size of the loop. Here, however, still a different form is
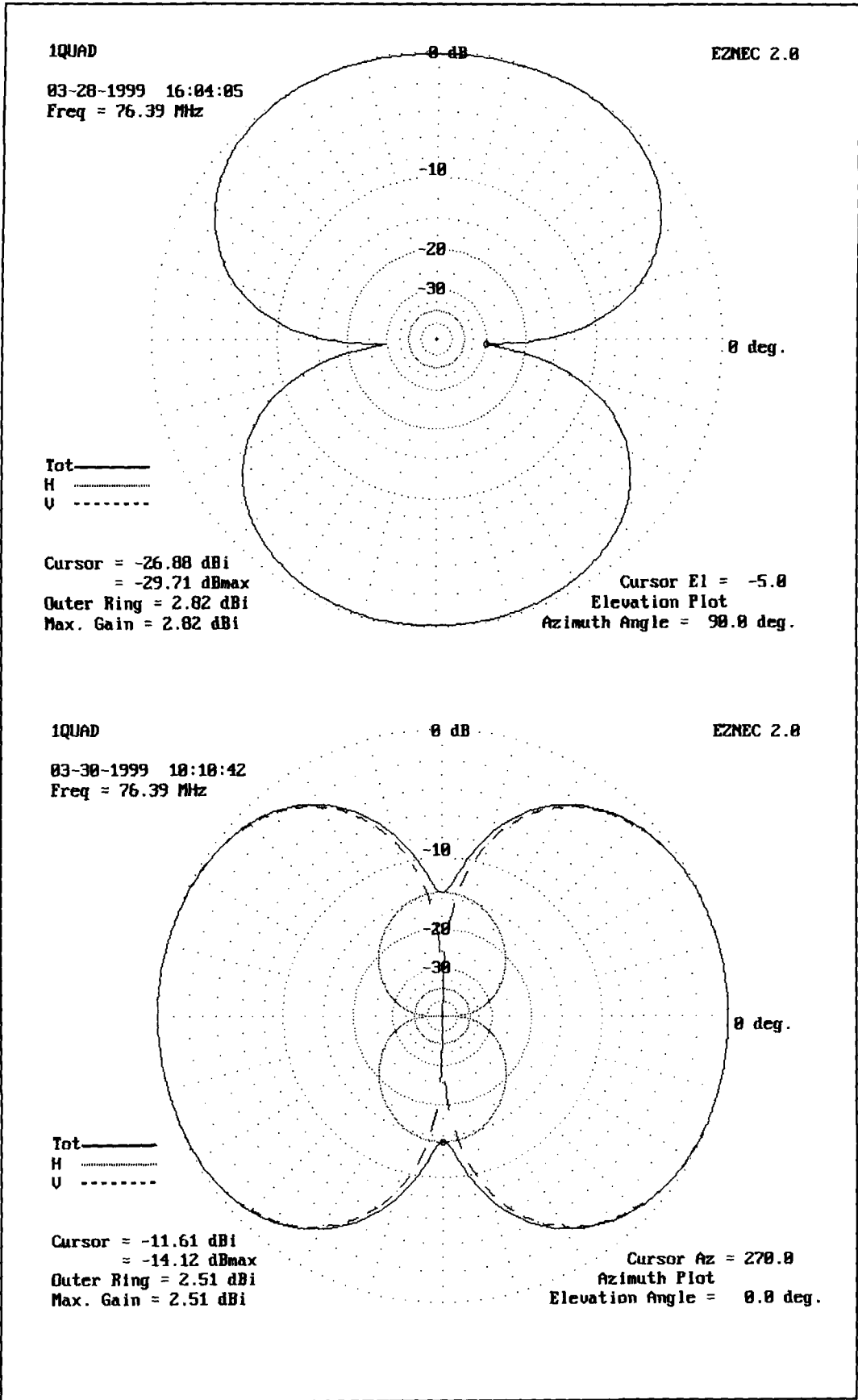
Figure 18. Quad radiation pattern, near second resonance.
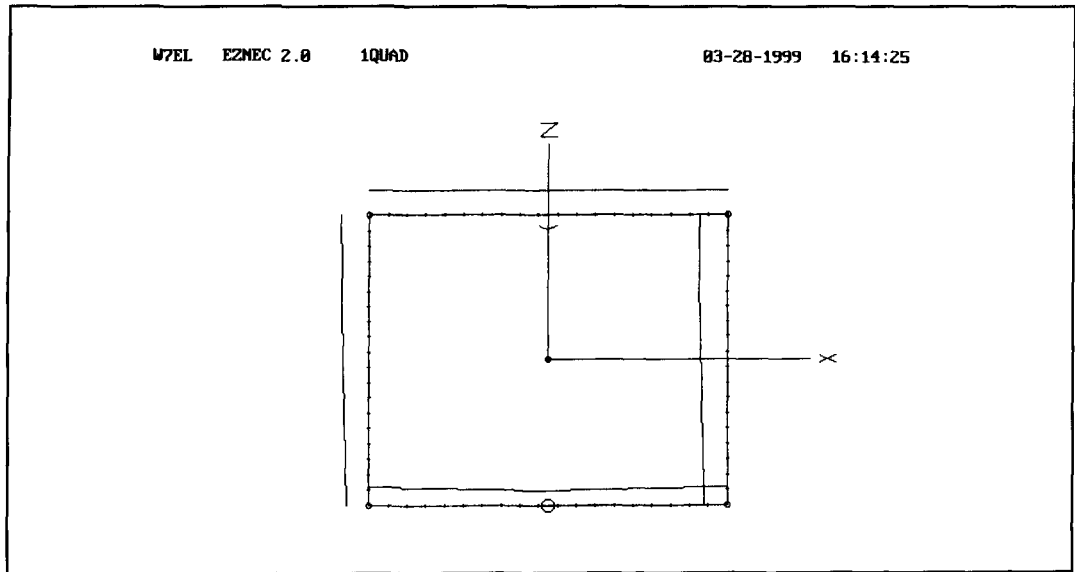
Figure 19. Current on loop well below parallel resonance.

also used. Define a reference frequency for a loop as 300/P, where P is the perimeter of the loop in meters. Now kb is the ratio, applied-frequency/reference-frequency. This is not the resonant frequency of the loop, which is a function of kb and the thickness factor. The resonant frequency will be denoted as fo.

Note on accuracy

The reference frequency uses the constant 300 instead of the more exact values of 299.8, or 299.7925. The greatest accuracy that can be expected in calculation is around one percent, and measurement errors are probably still larger. There is no reason to show more than three or even two significant figures. And, because of the likelihood of measurement errors, pattern calculations are usually shown here only to the nearest dB. This coarseness sometimes appears as bumps in the plots, but these are easily smoothed by the eye.

Quad loop drive impedance

The drive impedance of the family of quad loops is shown in **Table 1** for the range of kb from 0.05 to 2.0, and for thickness factors covering tubing to very fine wire. The reference frequency used for calculation was 37.5 MHz, for a loop with corners at ±1 meter, or 8 meters perimeter. The calculations are for copper.

Usually such tables are given as admittances, but impedance seems simpler for practical use. The conversion is g = 1/z, where g and z are vector quantities.

The impedance values are plotted in **Figure 5** for very fine wire. However, in one respect the plot is misleading. Both drive resistance and reactance change very rapidly around the first parallel resonant point kb~0.5. The step interval is too great: both the positive and negative peaks should be the same size. The second parallel resonance is adequately defined. The first series resonant point, kb~1.0 is easily seen at drive reactance equals zero.

Changing the step size to 0.5 MHz improves definition, as shown in **Figure 6**. This is sufficient to allow an estimate of the antenna Q, defined as F/F50 where F50 is the half intensity width. The value is about 11, indicating that the loop is fairly low Q. Still, smaller steps would be necessary for full definition.

Figure 7 is the impedance plot for a low thickness factor. The change at the first parallel resonant point is not great. The first series resonant point has moved upward appreciably, and the second parallel resonance has become broader, with less change in reactance. Still smaller values of thickness factor would show that the second parallel and higher resonances disappear: the quad becomes very broad-band.

Quad loop gain

The calculated gain for a quad loop is shown in **Figure 8** for kb up to 2.0. Two values are shown, the gain at right angles to the loop, and that in the plane of the loop. Note that there is gain above isotropic in some direction for all values of kb above first parallel resonance. At lower frequencies, both values become small.
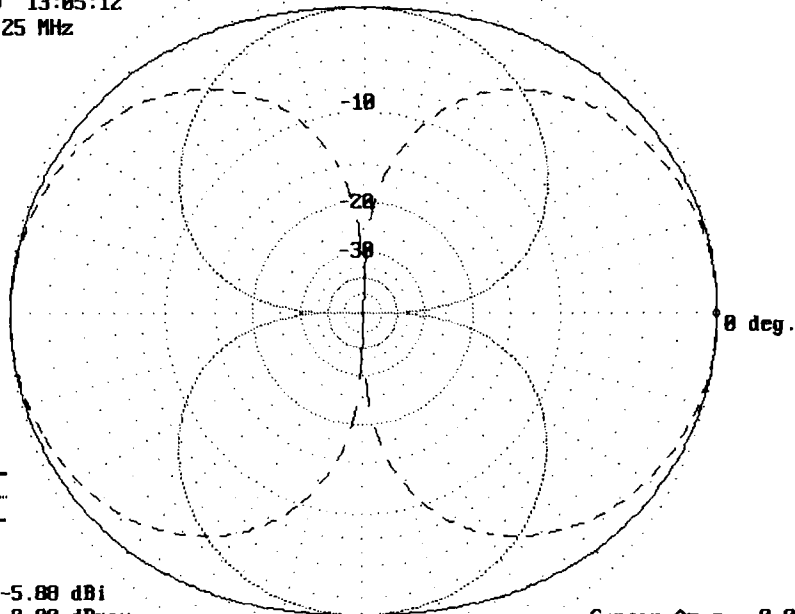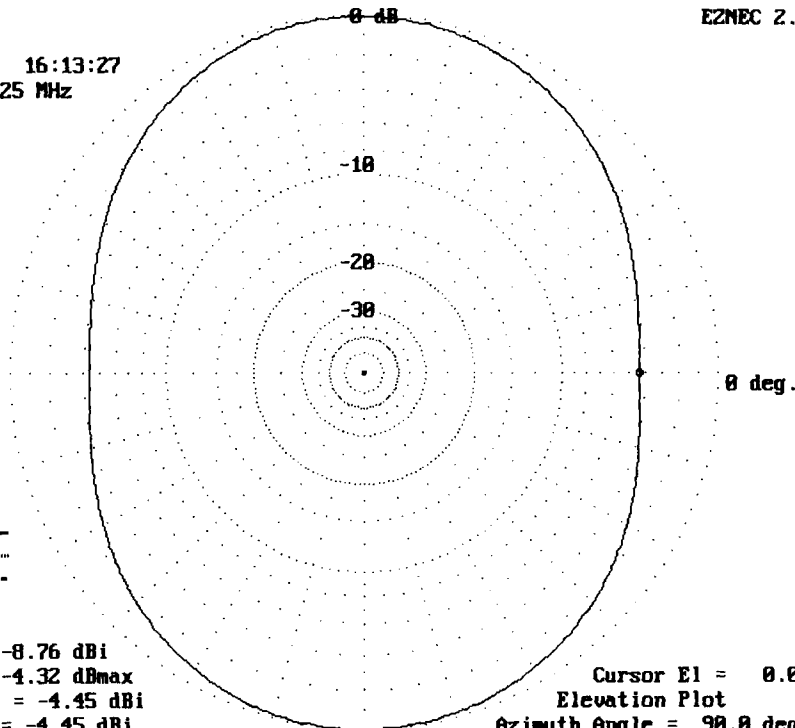
Figure 20. Radiation pattern, well below parallel resonance.
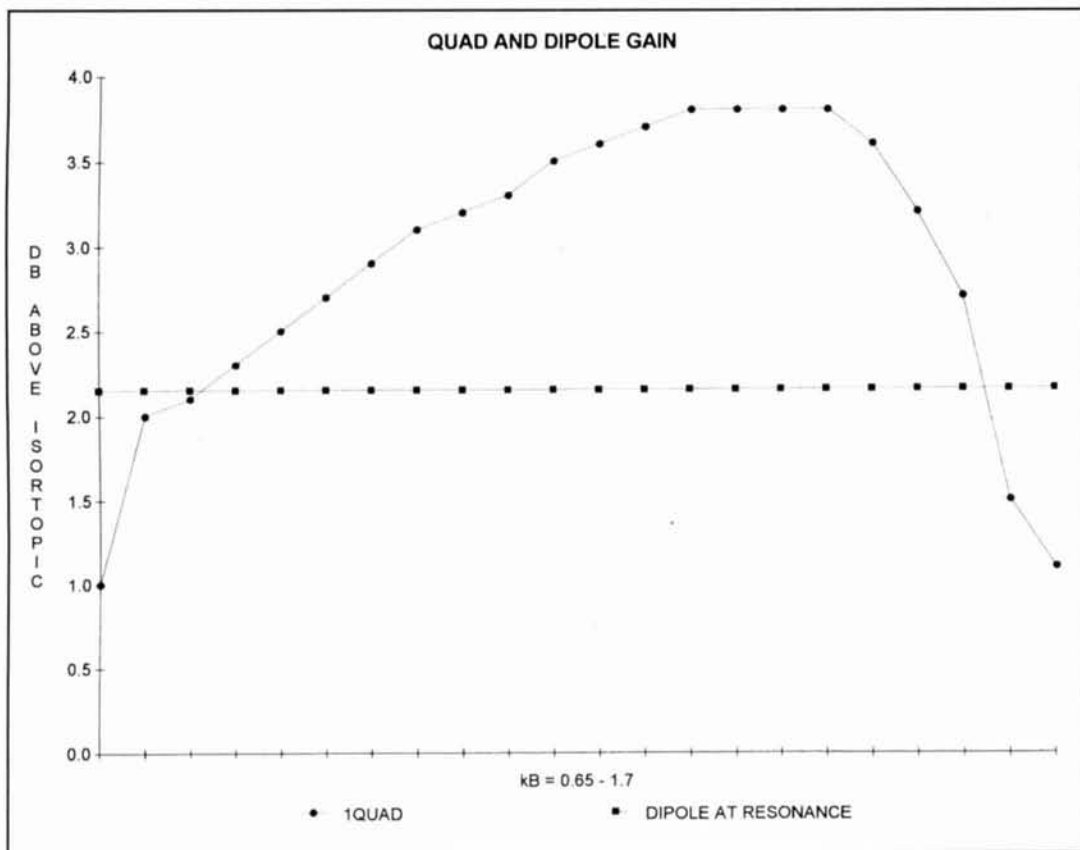
Figure 21. The quad as a broadband radiator.

Operation near first series resonance

The majority of quad operation is at or close to the first series resonance point, kb~1. **Figure 9** visualizes the current on the loop for kb = 1 (magnitude only). The fields due to the top and bottom sections add, for on-axis gain. Because there is a phase reversal in the other sides (usually the vertical ones), their on-axis fields cancel. Top and bottom wires have no field in the plane of the loop, but the side wires do because of their spacing.

The current is shown in greater detail in **Figure 10**. There is a standing wave around the loop. However, the current never goes to zero on the loop sides. The phase approximates a square shift, in phase with the drive for the lower half-wave of wire, and out of phase for the upper. The rate of change varies with thickness factor.

The pattern developed by wire geometry and current for the exact resonant point is shown in **Figure 11**, for bottom feed. The azimuth pattern at a) has a beamwidth of 90 degrees. However, the side lobe radiation is not zero as it would be for a dipole, due to the currents on the sides, which produce a small vertically polarized component—in this case, about 25 dB below maximum. Many quad users hold that they have excellent performance under poor conditions: this vertically polarized component is an easily demonstrated difference from dipole elements, but its importance is still not clear. It may be providing a type of signal arrival diversity.

Effects of wire thickness

The effect of wire size on quad gain is shown in **Figure 12**. Two effects appear. The gain increases with thick resonant wires, noticeably so for tubing sizes. This is due to the fact that the resonant frequency moves as the thickness factor becomes smaller, so a larger loop is needed for resonance. The other effect is a loss factor, which becomes effective for thin wires of real-world conductivity. For copper, the loss can be neglected for factors less than about 3000, larger than #22 AWG at 37.5 MHz. I find #18 satisfactory at 20 meters, but #12–14 would be better at 40 meters.

This can be regarded as a consequence of a lower radiating area.

RECTANGULAR QUADS
RESISTANCE VS SHAPE FACTOR

KB=0.85 - 1.1, STEP .05

SF=.125   SF=.25   SF=.375
SF=.5   SF=.625   SF=.75
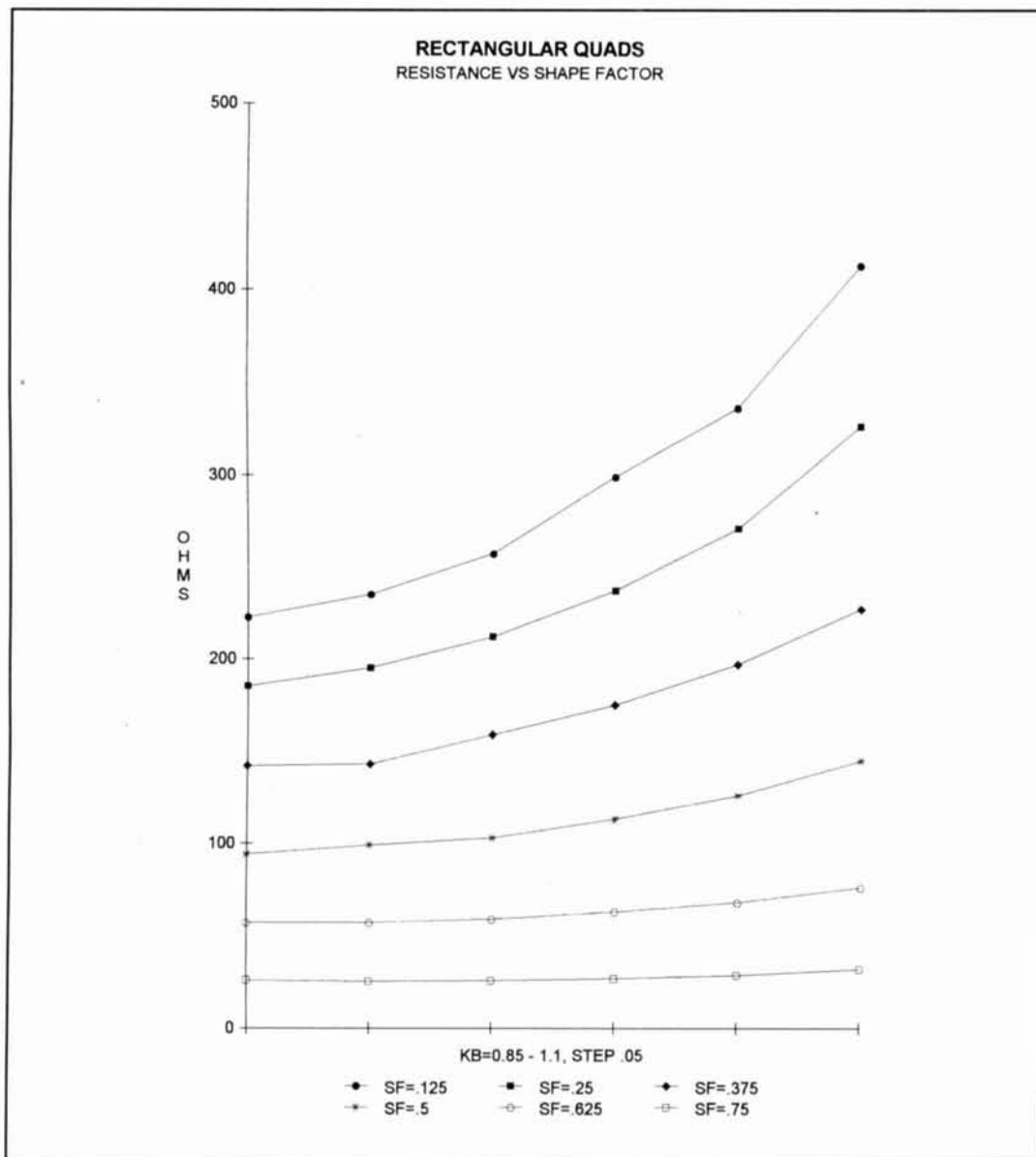
Figure 22. Resistance of rectangular quads near resonance.

The resonant frequency as a function of thickness factor is shown in **Figure 13**. An isolated resonant bare wire quad loop is always more than a wavelength in perimeter.

It is sometimes convenient to calculate the resonant frequency by a relation. Let $X = \log(p/2a)$; log to the base 10 of the thickness factor. To good accuracy for thick conductors or tubing, the resonant $fo/fr = 1.461 - 0.1886 X + 0.021 X^2$.

The variation in drive resistance is shown in **Figure 14**. Again, two effects appear. Drive resistance at resonance increases as the thickness factor becomes smaller. This is largely a result of the increase in loop size needed to produce resonance. Also, the drive resistance becomes larger with very fine wires—the result

of wire loss. Again, wire larger than a thickness factor of 3000 shows small effect. Over the range from 10,000 to 1000, the resonant drive resistance is nearly constant at 128 ohms.

The drive resistance does change with frequency, as shown in **Figure 15** for frequencies close to resonance. For 20 meters, the band width is only 0.025 f/fr units wide, so the resistance change is negligible. Even on 10 meters, with about three times the bandwidth, the change can be ignored.

The change in drive reactance is more important, as shown in **Figure 16**. The intersection of the lines with the zero reactance axis marks the resonant frequencies for the various wire sizes. Except for the extremes in wire size, the curves can be expressed in point-slope form, where the
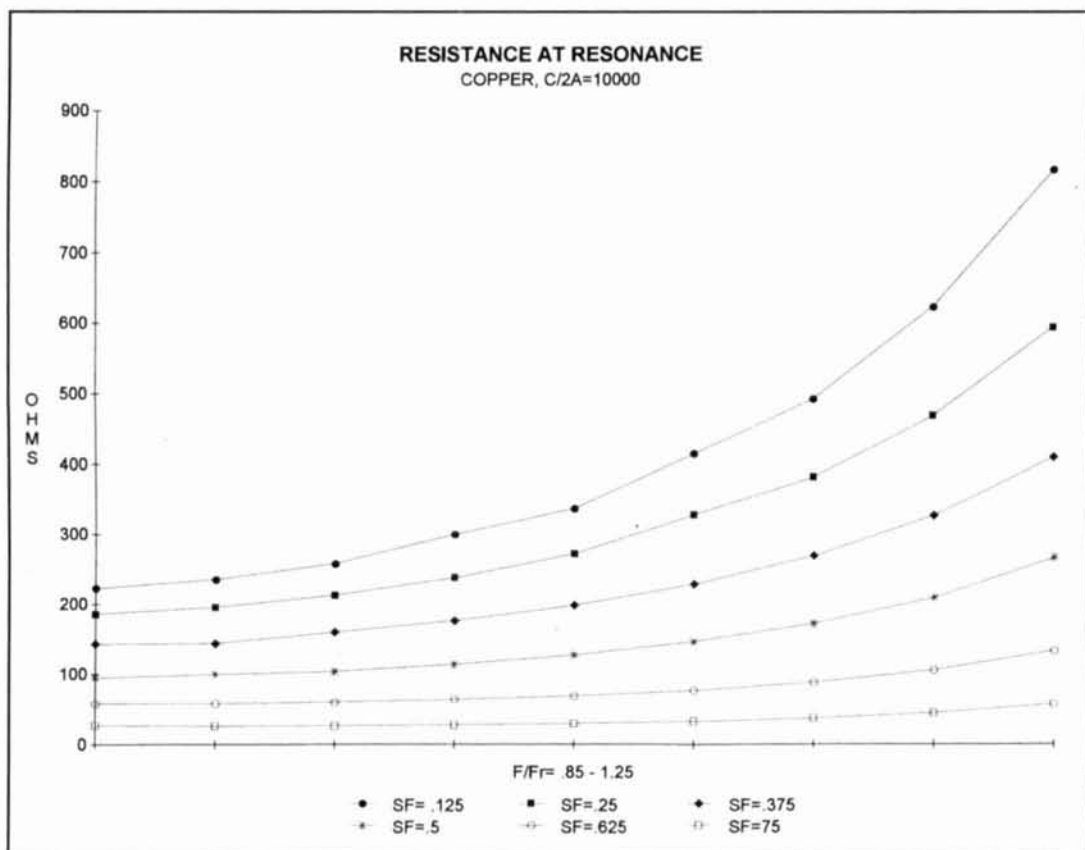
Figure 23. Resistance at resonance versus shape factor.

point is the value at which all thicknesses show the same reactance, essentially -140 ohms for F = Fr = 1.0. These curves are also important for further work. For example, they show the change in dimension needed to remove reactance from the drive point.

Quad performance at other frequencies

Figure 17 shows the magnitude of the currents at the second series-resonant frequency, kb~2.0. If the alternating current phases are observed, it is seen that the horizontally polarized radiation from the top and bottom wires nearly cancels on the loop axis, as in **Figure 18**; however, the cancellation is not complete because the currents are not quite equal. The vertically polarized component is large and also shows on the elevation angle plot at b).

Well below the resonant frequency, the current is nearly constant over all parts of the antenna, as shown in **Figure 19**. This condition is often used as the definition of a small loop, such as the type used in direction finding. The total field azimuth plot is nearly a circle, as in **Figure 20**, this being the sum of two figure-8

patterns of the different polarizations. The elevation plot b) is horizontally polarized at right angles to the loop plane.

Figure 8 shows that maximum sensitivity to vertical polarization is at kb~0.4. Some rejection of the effect of horizontal polarization can be secured by operation at kb~0.15. (Other types of antennas exist for direction finding.)

Quad loops as wide-band antennas

Reference has been made to the relative constancy of quad loop drive resistance, and the relatively low rate of change of reactance with frequency change. These are indications that a quad loop can serve as a good wide-band antenna. **Figure 21** indicates the useful range: the on-axis loop gain exceeds that of a dipole from kb = 0.76 to kb = 1.64. This means, for example, that a 14-MHz loop is a good radiator at 10 MHz and also at 18 and 21 MHz.

There are, of course, penalties for such use. Drive resistance and especially reactance can vary quite widely, as shown in **Table 1** and **Figure 5**. Low-loss feedline and a good antenna tuner takes care of such problems.

Table 2. Gain and impedance data for quad variations.

FILE		1LLLQ SF=.125		1LLQ SF=.25		1LQ SF=.375		1HQ SF=.625		1HHQ SF=.75		1HHHQ SF=.875		1DIAM		1OCTO	
SHAPE FACTOR																	
kB	F	R	X	R	X	R	X	R	X	R	X	R	X	R	X	R	X
0.85	31.875	222.4	-492	185	-547	142	-589	57	-597	25.9	-558	8	-483	92	-610	101	-554
0.9	33.75	234.9	-291	195	-361	143	-410	57	-427	25.4	-392	7.7	-328	96	-437	105	-390
0.95	35.625	257.1	-116	212	-200	159	-255	59	-277	25.7	-245	7.6	-188	103	-286	113	-245
1	37.5	298.9	45	237	-52	175	-114	63	-139	26.8	-107.5	8	-56	113	-146	125	-108
1.05	39.375	335.8	200	271	88	197	21	68	-4.2	28.7	27.9	8.5	76	128	-11	141	26
1.1	41.25	412.8	356	326	227	227	155	76	133	31.9	168	9.5	213	148	125	164	163
1.15	43.125	489.9	518	380	369	268	292	88	279	36.6	320	11.2	366	178	263	196	310
1.2	45	620.2	688	466	518	325	439	105	443	44.1	496	13.9	544	219	425	243	476
1.25	46.875	815.3	869	591	677	407	601	132	640	56.5	711	18.6	769	280	606	316	672

kB	F	MAXGAIN	VG	MG	VG	MG	VG	MG	VG	MG	VG	MG	VG	MG	VG	MG	VG
0.85	31.875	2	-25	2.1	-19	2.2	-16	2.8	-12	3.1	-12	2.5	-12	2.4	-13	2.7	-14
0.9	33.75	2.1	-28	2.2	-22	2.4	-18	3.1	-15	3.5	-14	2.9	-15	2.6	-15	2.9	-16
0.95	35.625	2.1	-30	2.2	-24	2.5	-21	3.4	-18	3.8	-18	3.3	-19	2.8	-17	3.1	-20
1	37.5	2.2	-30	2.3	-26	2.6	-24	3.6	-23	4.1	-25	3.6	-30	3	-19	3.3	-24
1.05	39.375	2.2	-27	2.4	-25	2.7	-25	3.8	30	4.4	-34	3.8	-27	3.1	-20	3.5	-25
1.1	41.25	2.3	-25	2.4	-22	2.9	-22	4	-23	4.6	-21	4	-18	3.2	-19	3.6	-21
1.15	43.125	2.3	-23	2.5	-20	2.9	-19	4.2	-18	4.8	-16	4.1	-14	3.4	-18	3.8	-18
1.2	45	2.3	-21	2.5	-17	3	-16	4.3	-14	4.9	-13	4.1	-11	3.5	-16	3.9	-15
1.25	46.875	2.4	-20	2.6	-16	3	-14	4.5	-11	5	-10	3.9	-9	3.6	-14	4	-13

	Fres	R	Fres	R	..	R	Fres	R	Fre	R	Fres	R	Fres	R	Fres	R
COPPER	36.975	279	38.2	248		193	39.435	68.3	38.99	28.3	38.3	8.1	39.53	130	39.01	137
IDEAL	37	277	38.23	246		191	39.47	68.7	39.02	25.6	38.33	5.5	39.56	127	39.04	135

1OCTO P/2A	Fres	R copper	LOG P/2A
100	42.91	156.9	2
200	41.28	145	2.30103
500	40.29	140	2.69897
1000	39.84	138	3
3000	39.37	138	3.4771213
10000	39.01	137	4
30000	38.74	142	4.4771213
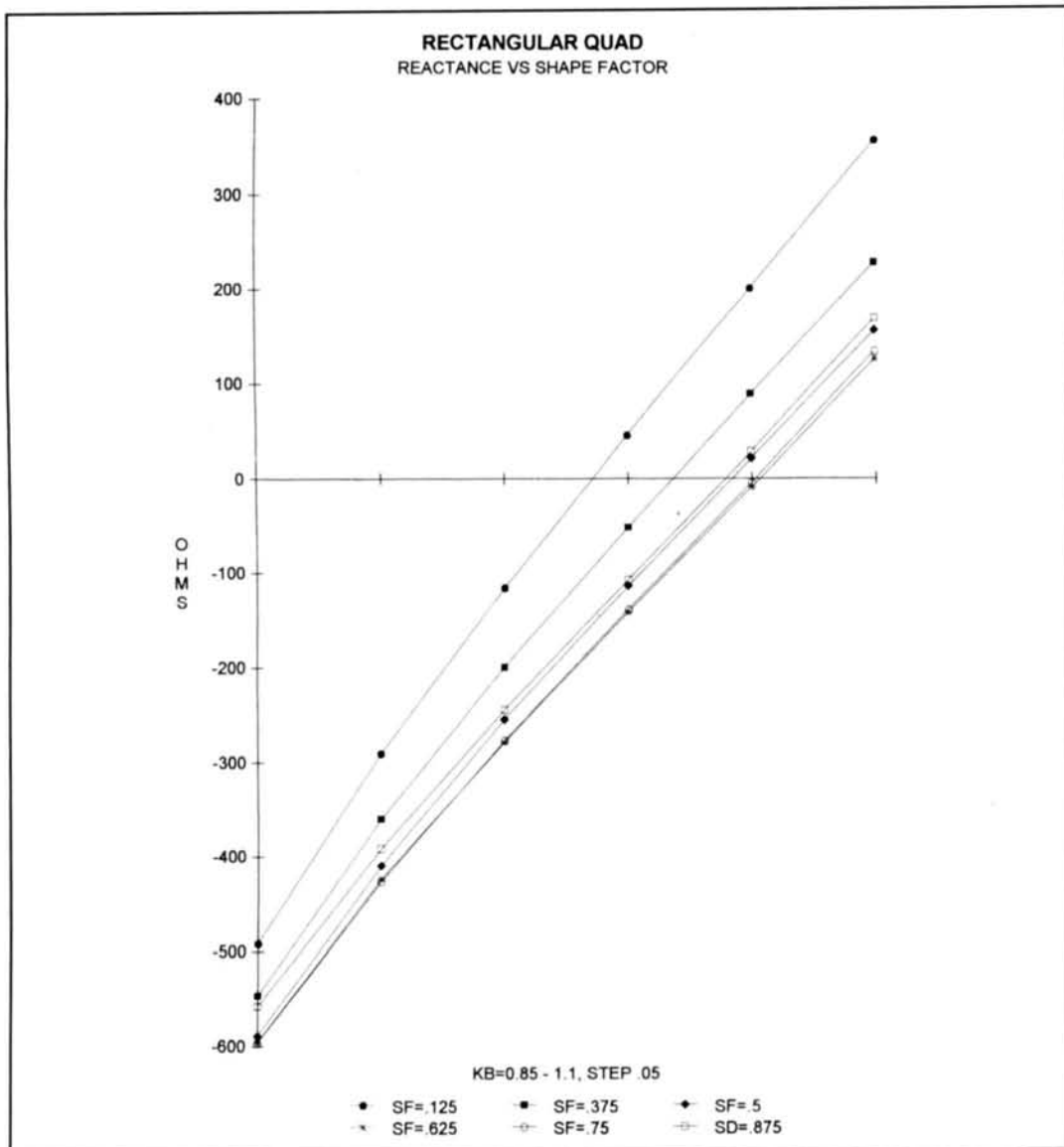
Figure 24. Reactance of rectangular quads near resonance.
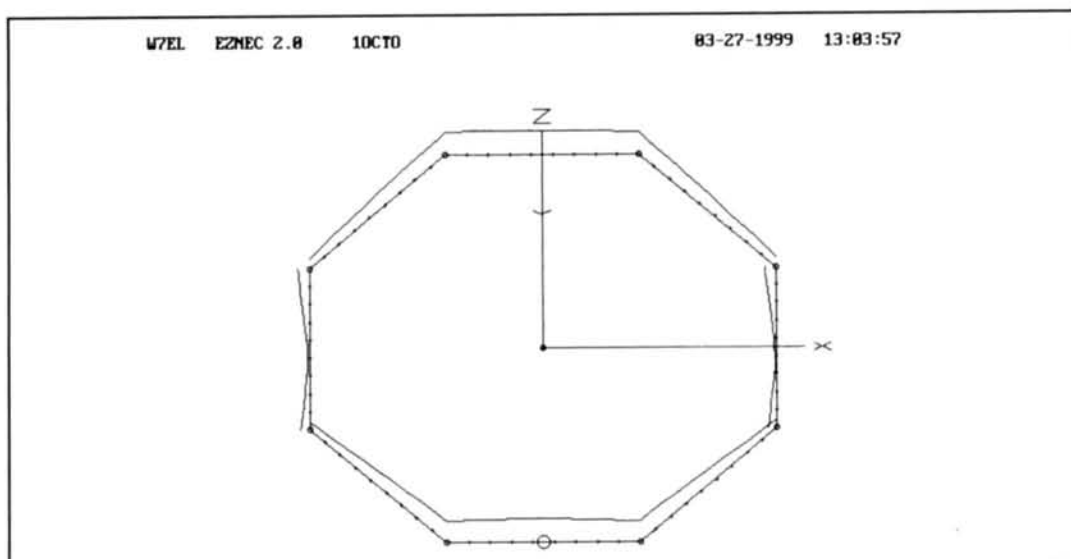
Figure 25. Octagon antenna near series resonance.

10CTO   EZNEC 2.0

03-27-1999 13:02:06
Freq = 37.5 MHz

Tot————
H   ·····················
U   --------

Cursor = -20.80 dBi
       = -24.11 dBmax
Outer Ring = 3.31 dBi
Max. Gain = 3.31 dBi

Gain: 3.31 dBi
Angle: 90 deg
F/Side: 24.11 dB
Bmwidth: 83.6 deg
-3: 48.2, 131.8 deg
Slobe: 3.31 dBi
Angle: 270 deg
F/Slobe: 0.00 dB

0 deg.

Cursor Az =   0.0
Azimuth Plot
Elevation Angle =   0.0 deg.

10CTO   EZNEC 2.0

03-28-1999 16:17:03
Freq = 37.5 MHz

Tot————
H   ·····················
U   --------

Cursor = -0.55 dBi
       = 0.00 dBmax
Outer Ring = -0.55 dBi
Max. Gain = -0.55 dBi

0 deg.

Cursor Az =   0.0
Azimuth Plot
Elevation Angle = 90.0 deg.

Figure 26. Pattern of octagon antenna at series resonance.
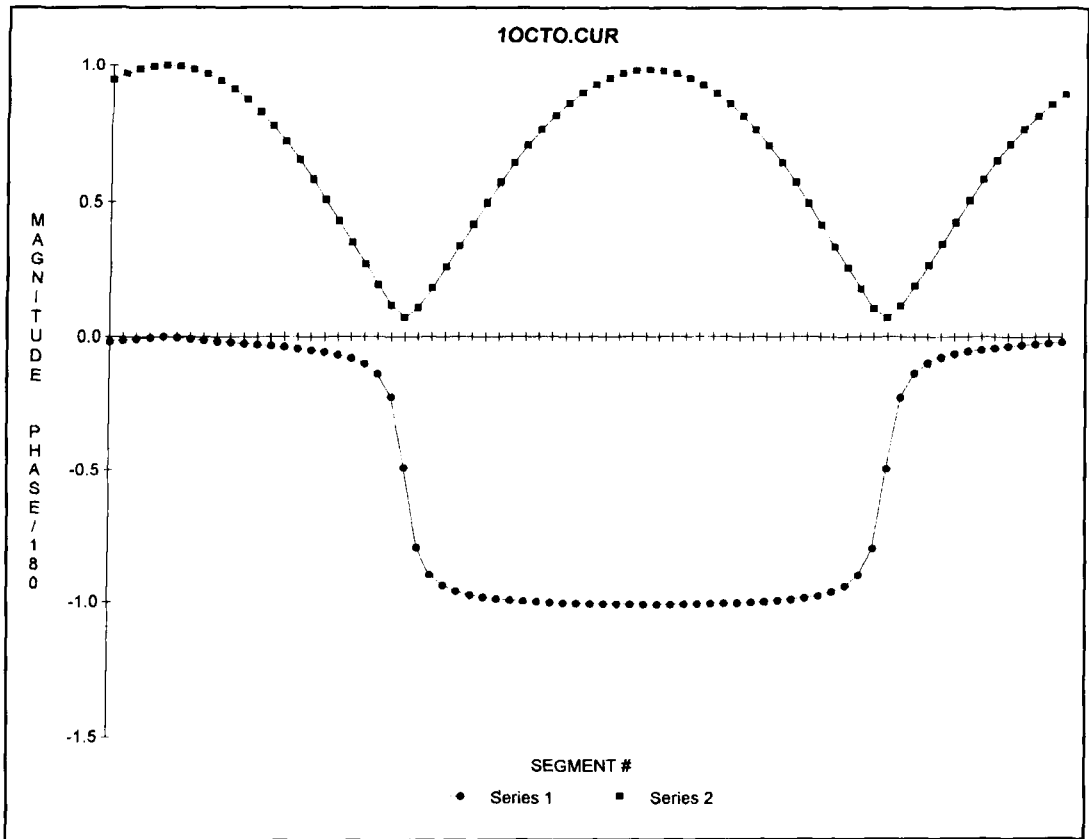
Figure 27. Current on octagon antenna, 0.8-mm wire.
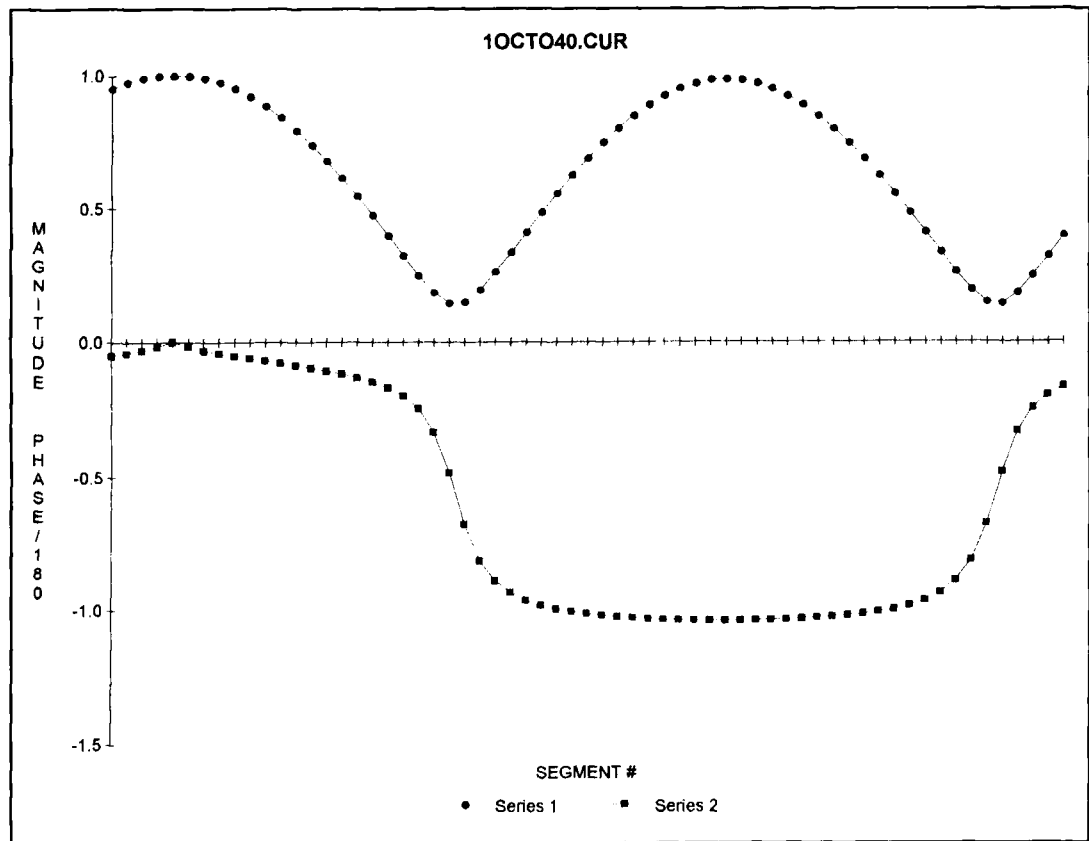
Figure 28. Current on octagon antenna, 40-mm wire.
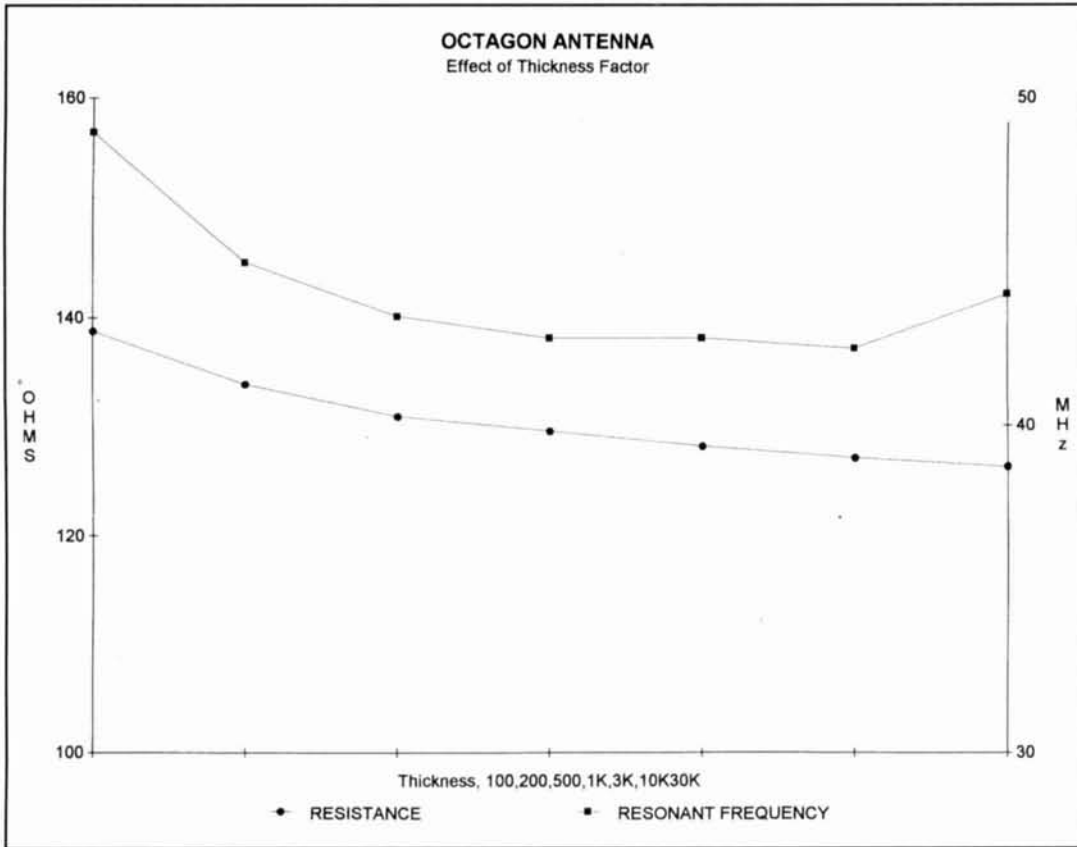
Figure 29. Effect of wire thickness on octagon antenna.
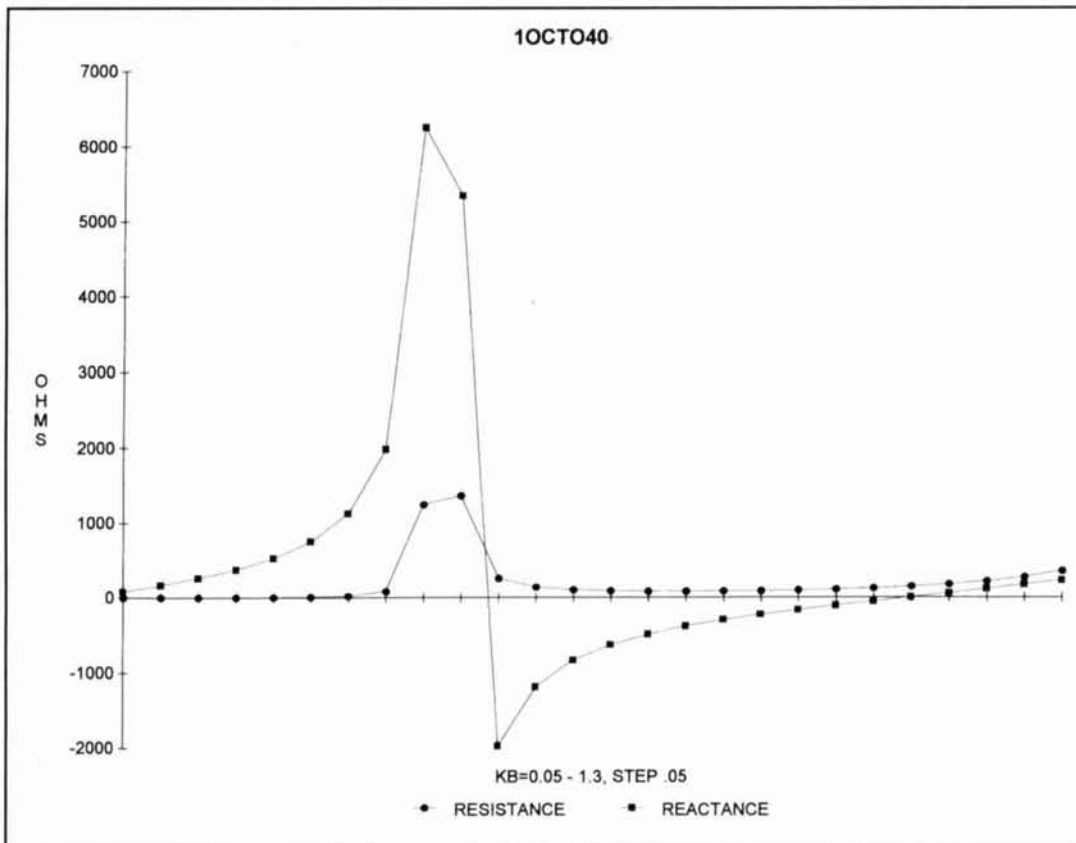
Figure 30. Impedance of octagon at "Army loop" frequencies.
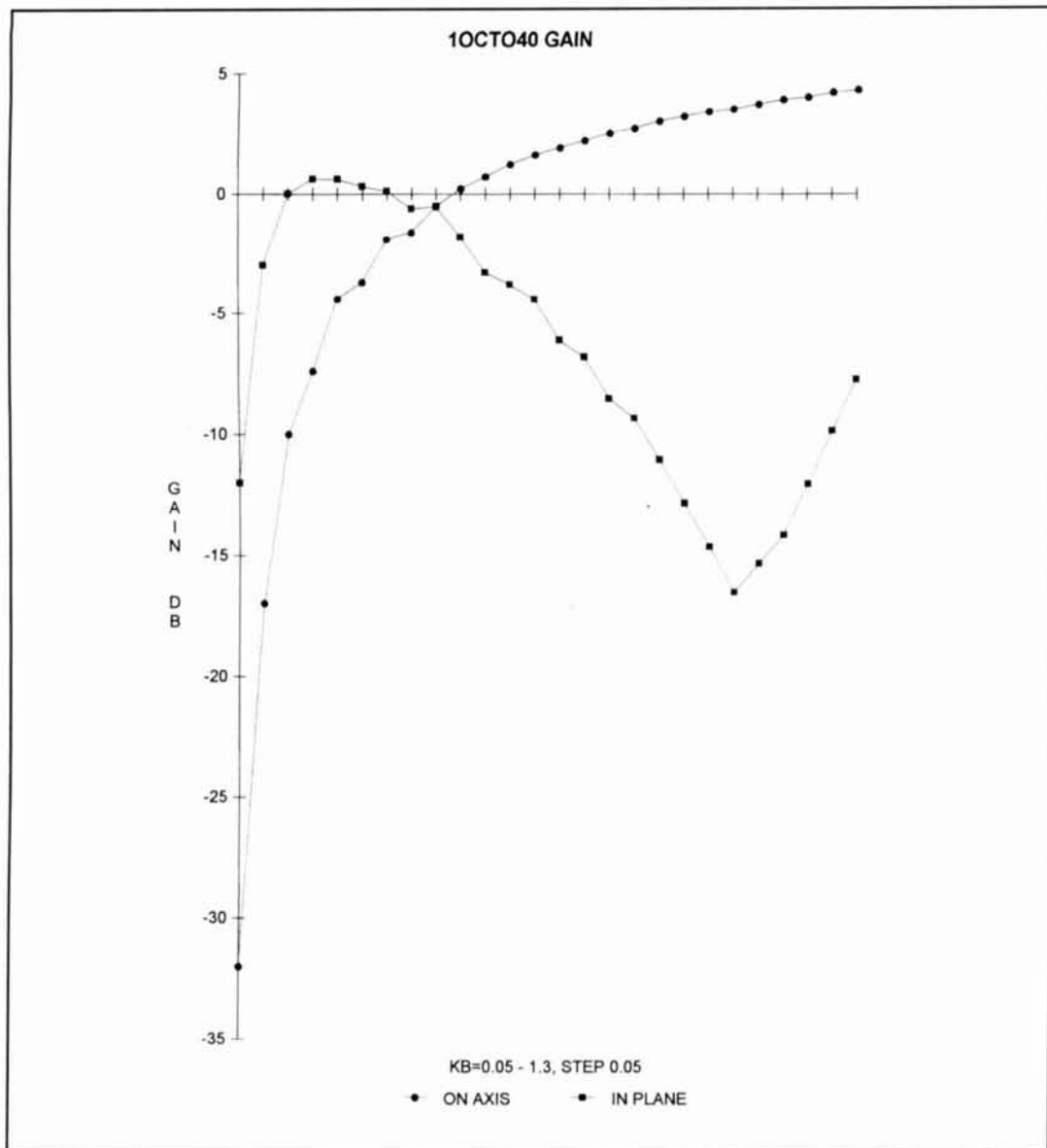
Figure 31. Gain of octagon at "Army loop" frequencies.

Figure 21 also shows another possibility. Almost 1 dB more gain can be secured by operating a radiator above the resonant frequency. Again the problems of varying load resistance and reactance appear, but again low-loss line and an antenna tuner handle these. Later we will see that making quad radiators and reflectors the same physical size is a practical way to exploit this possibility.

Other loop shapes

The square loop is only one of the family of loops, and, in principle, any geometric figure having area can serve as a quad element. However, for mechanical reasons, such as the num-ber of support arms needed, relatively few shapes are commonly used. Some will be considered here, and a few others will be touched upon later as less common antenna problems are considered.

The rectangular loop

The square loop is really a member of the family whose shape is a rectangle. These include, at one extreme, a folded dipole, often used as a starting point for quad visualization. At the other extreme, the loop is so narrow that it becomes a shorted transmission line. It is the rectangles between these extremes that are of interest as quad possibilities.
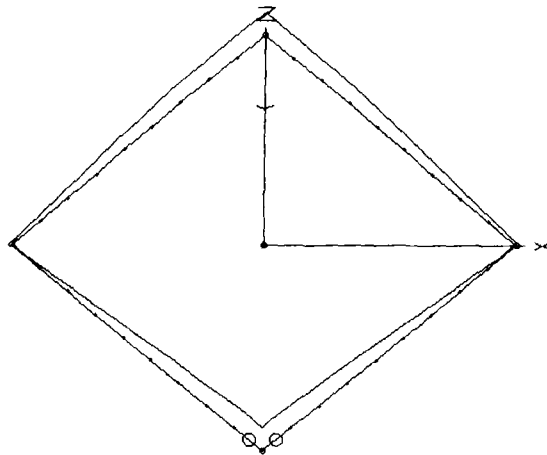
Figure 32. Diamond antenna at series resonance.

Figure 3 includes the definition of shape factor, which is 0.5 for a square loop, almost zero for a folded dipole, and almost unity for the transmission line. This quantity is a controlling factor for the family.

Table 2, at the top, shows the feed resistance and reactance for the rectangular loops, for the range kb = 0.85 to 1.25, and for shape factors from 0.125 to 0.875 (the octagon and diamond entries will be covered later). The center part of the table shows the gain: in this range the maximum gain MG is the horizontally polarized on-axis gain. VG is the vertically polarized gain in the plane of the loop, in dB relative to isotropic. Just below these values is a list of resonant frequencies at resonance drive resistances for the shapes, for ideal conductors and for copper.

Figure 22 shows the drive resistance in graphical form. For the "squashed down" loops with SF = 0.125, the drive resistance is around the 300-ohm value of a folded dipole. For a large shape factor, the drive resistance moves towards zero. **Figure 23** replots the data to show explicitly the effect of shape factor: both show the increase in drive resistance as frequency increases for frequencies not far from resonance.

Figure 24 shows the variation of reactance with frequency for a number of shape factors. The curves are essentially parallel to the line for the square quad, but large values of SF have a different effect from the low ones. This reflects the difference between a folded dipole and a shorted line.

As shown in **Table 2**, there is not an enormous change in gain over the range of SF used

here. If practical considerations indicate use of a partly squashed or narrowed shape, performance will not change much. Shape factors of 0.6 to 0.75 operated above resonance can give more gain than a conventional square. Narrowing the shape is also a way to control the drive resistance, if feed line matching is a goal.

The octagon antenna

Because of the extra support needed, octagon shapes are rare in wire antennas. A reason for looking at them is that they are close to the circular shapes common in theoretical analysis. Small octagons built of tubing large enough to be low loss and self supporting are more common, such as the "Army loop."

Calculated data for the octagon loops is given in the bottom part of **Table 2**. The shape of the antenna is shown in **Figure 25**, which also shows the current distribution. Note that the calculations are based on nine segments per side, which doubles the number of segments per wavelength as compared to a square loop.

The pattern for the octagon is much the same as for the square loop, which can be seen by comparing **Figure 26** (octagon) with **Figure 11** (square loop). The major difference is the azimuth plot, a circle instead of an oval. The maximum difference is about 4 dB. The difference is due to the larger product, current-times-degrees in the vertical direction. However the current per segment is almost exactly the same in octagon and square loops, as can be seen by comparing **Figures 27** and **10**. Note that the

1DIAM   0 dB   EZNEC 2.0

03-27-1999 12:32:14
Freq = 37.5 MHz

                                          -10

                                          -20

                                          -30

                                                                            0 deg.

                                        Gain: 2.95 dBi
                                        Angle: 90 deg
                                        F/Side: 24.23 dB
                                        Bmwidth: 83.8 deg
                                        -3: 48.1, 131.9 deg
Tot————   Slobe: 2.95 dBi
H ················   Angle: 270 deg
U - - - - - - -   F/Slobe: 0.00 dB

Cursor = -21.28 dBi
       = -24.23 dBmax   Cursor Az =   0.0
Outer Ring = 2.95 dBi   Azimuth Plot
Max. Gain = 2.95 dBi   Elevation Angle =   0.0 deg.

1DIAM   0 dB   EZNEC 2.0

04-05-1999 13:45:59
Freq = 37.5 MHz

                                          -10

                                          -20

                                          -30

                                                                            0 deg.

Tot————
H ················
U - - - - - - -

Cursor =   2.95 dBi
       =   0.00 dBmax   Cursor El =   0.0
Outer Ring = 2.95 dBi   Elevation Plot
Max. Gain = 2.95 dBi   Azimuth Angle =   90.0 deg.

Figure 33. Pattern of diamond antenna at series resonance.
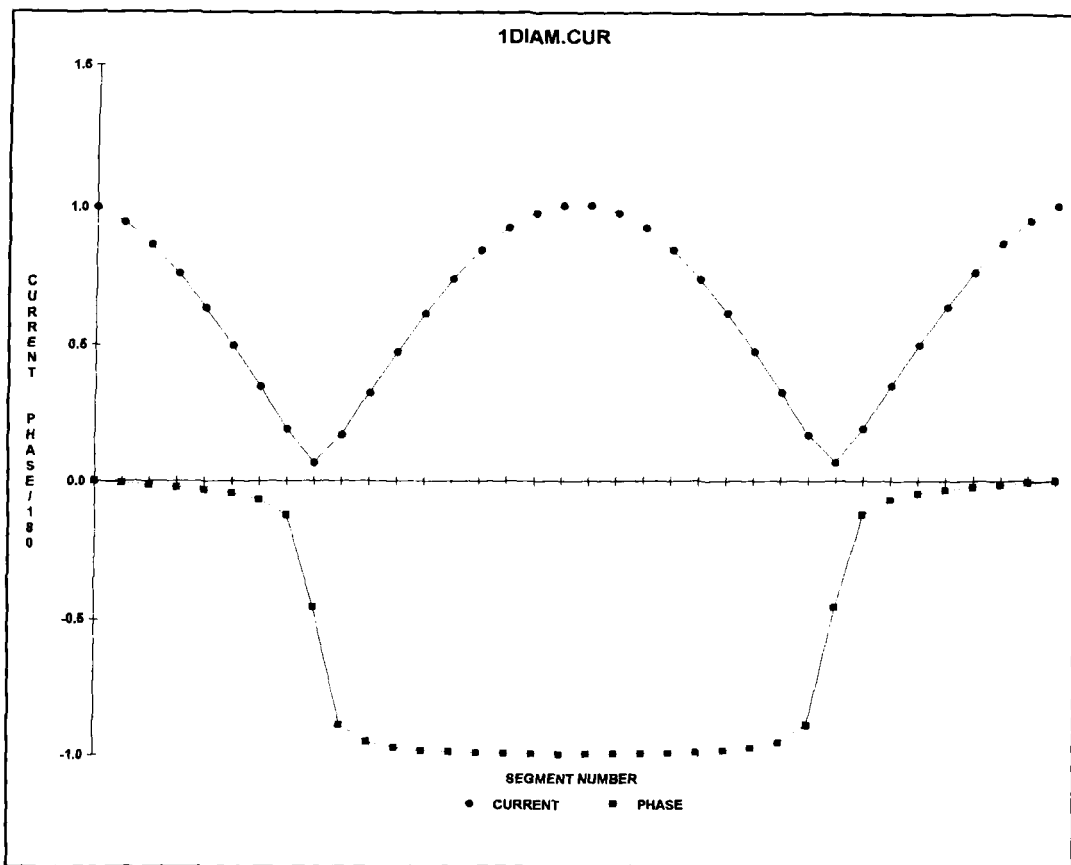
Figure 34. Current on diamond antenna at series resonance.

segmentation scheme changes the drive point on the plots.

The Army loop

The use of tubing instead of wire, as in the Army loop, changes the current distribution somewhat, as seen by comparing **Figures 27** and **28**. The minimum current is larger, and the current and phase changes are slower. This is characteristic of a lower Q system. The lower Q is evident in **Figure 29**, the variation of drive resistance and reactance with frequency. Reactance remains high, but the resistance is markedly lower at the first parallel resonant point. The change due to conductor size is less marked at the first series resonant point, shown in **Figure 30**. This also shows the change in resonant frequency.

The Army loop is a small antenna, using a tuning capacitor to remove reactance. A result is that gain can be small, as already seen in small loops. **Figure 31** shows its variation. Maximum small loop gain occurs at kb = 0.2, but this often means too large an antenna. Better on-axis versus in-plane discrimination occurs with smaller sizes. For receiving antennas, the best choice requires consideration of

the expected noise levels: as long as the ambient noise (static) is greater than the circuit noise, there is no S/N penalty resulting from reducing antenna size. In transmitting, the key factors are the drive resistance as compared to the loss resistance. Practical size Army loops must use the best of low-loss construction.

The diamond loop

If a square loop is rotated 45 degrees, it takes the shape of a diamond. Because wire length and angles at joints remain the same, there will be relatively little change in characteristics. There will be some, however, because the two high-current points are further apart than in the square. The same is true of the two high-voltage points at the other corners. This is illustrated in **Figure 32**, which also illustrates a limitation of segment analysis in NEC. Excitation is assumed to be at the center of a segment. One way of handling this is to introduce a short segment, replacing the 90-degree angle. Another, illustrated here, is to assume that half of the current is introduced on each of the two segments on either side of the corner. Neither are exact, but the error is small to negligible.

1DIAM

03-09-1999 16:51:44
Freq = 78.06 MHz

0 dB

EZNEC 2.0

-10

-20

-30

0 deg.

Gain: 1.81 dBi
Angle: 135 deg
F/B: 0.20 dB
Bmwidth: 139 deg
-3: 110.5, 249.5 deg
Slobe: 1.81 dBi
Angle: 225 deg
F/Slobe: 0.00 dB

Tot ————
H ·············
V - - - - - -

Cursor = -16.03 dBi
       = -17.84 dBmax
Outer Ring = 1.81 dBi
Max. Gain = 1.81 dBi

Cursor Az = 270.0
Azimuth Plot
Elevation Angle =   0.0 deg.

1DIAM

04-05-1999 13:49:20
Freq = 78.06 MHz

0 dB

EZNEC 2.0

-10

-20

-30

0 deg.

Tot ————
H ·············
V - - - - - -

Cursor = -16.03 dBi
       = -18.55 dBmax
Outer Ring = 2.52 dBi
Max. Gain = 2.52 dBi

Cursor El =   0.0
Elevation Plot
Azimuth Angle = 90.0 deg.

Figure 35. Pattern of diamond antenna at second series resonance.
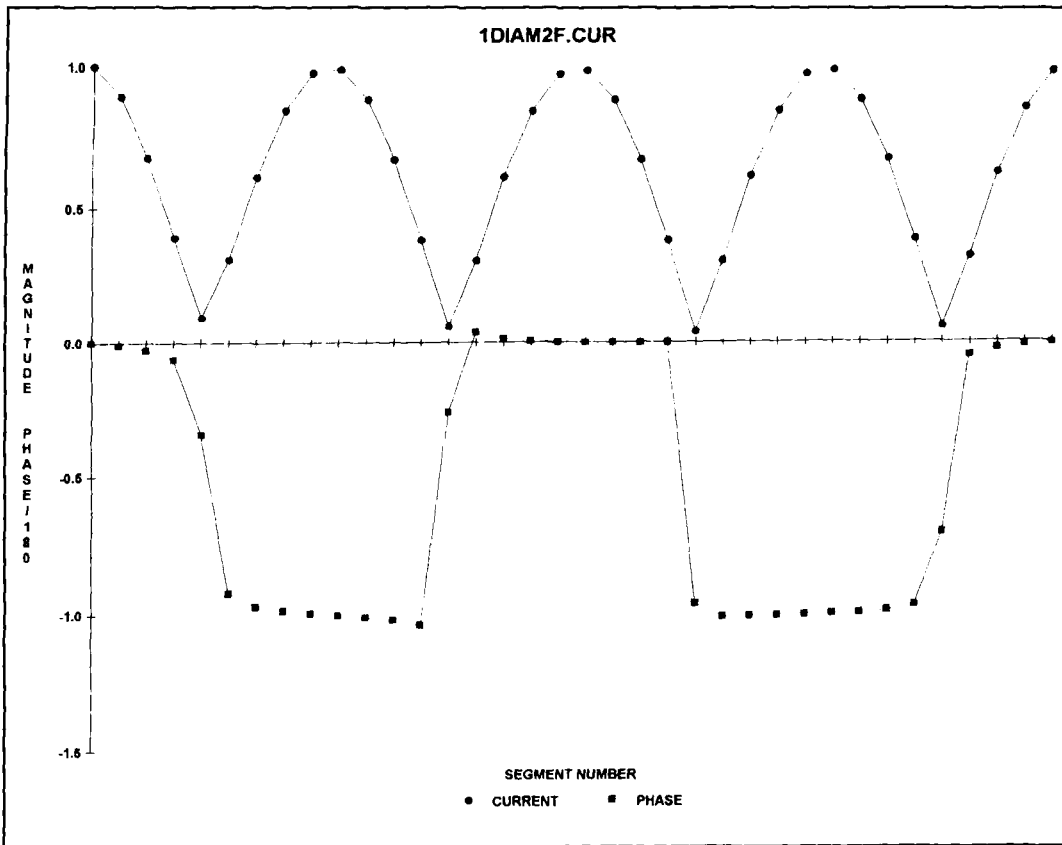
Figure 36. Current on diamond antenna at second series resonance.

The pattern of the diamond is shown in **Figure 33**. It is almost identical to that of the square loop, **Figure 11**. The same is true for the current distribution (compare **Figure 34** with **Figure 10**).

The pattern at the second series resonance, **Figure 35**, does show some variation, largely due to the 45-degree rotation. This figure also illustrates the distortion introduced by use of a single feed at a corner in NEC; it is especially noticeable in the vertical component. MININEC has the same problem, but it almost disappears in the thin wire program, which allows feed at the end of a segment.

Figure 36 illustrates the current distribution on a diamond.

Insulation on quad wires

There does not seem to be a real need for insulation on the wires of a quad loop. True, it reduces the corrosion in bad atmospheres; however, my experience with bare copperweld has not shown any problems in this respect— even with wires in use for five or six years in Florida, a few miles from the ocean. Both iron and aluminum corrode badly in less time.

Commercial versions of quads, especially 2-meter designs, often use insulated wire. If use of such wire is indicated, be aware that the data presented so far cannot be used directly, and that more calculations are needed to account for the presence of insulation. (The insulation thickness for enameled wire is too small to have an effect.)

The general effect of insulation can be seen by considering a wire segment as a lumped element. There is no effect on the inductance of the segment, assuming non-magnetic insulation, as normal. However, the capacitance does increase—the amount being determined by the insulation thickness and dielectric constant. Because the overall resonant frequency varies as the square root of $1/(LC)$, the insulation moves the resonant frequency down, affecting the entire reactance curve. Also, because resistance varies as the square root of (L/C), the drive resistance decreases.

Exact values for all variables presented so far may be determined with the thin wire and NEC4 programs. Here an approximate method[7] is used, which is based on the thin wire program and the assumption that operation is close to resonance.

INSULATED WIRE CORRECTION
EFFECT OF P-FACTOR
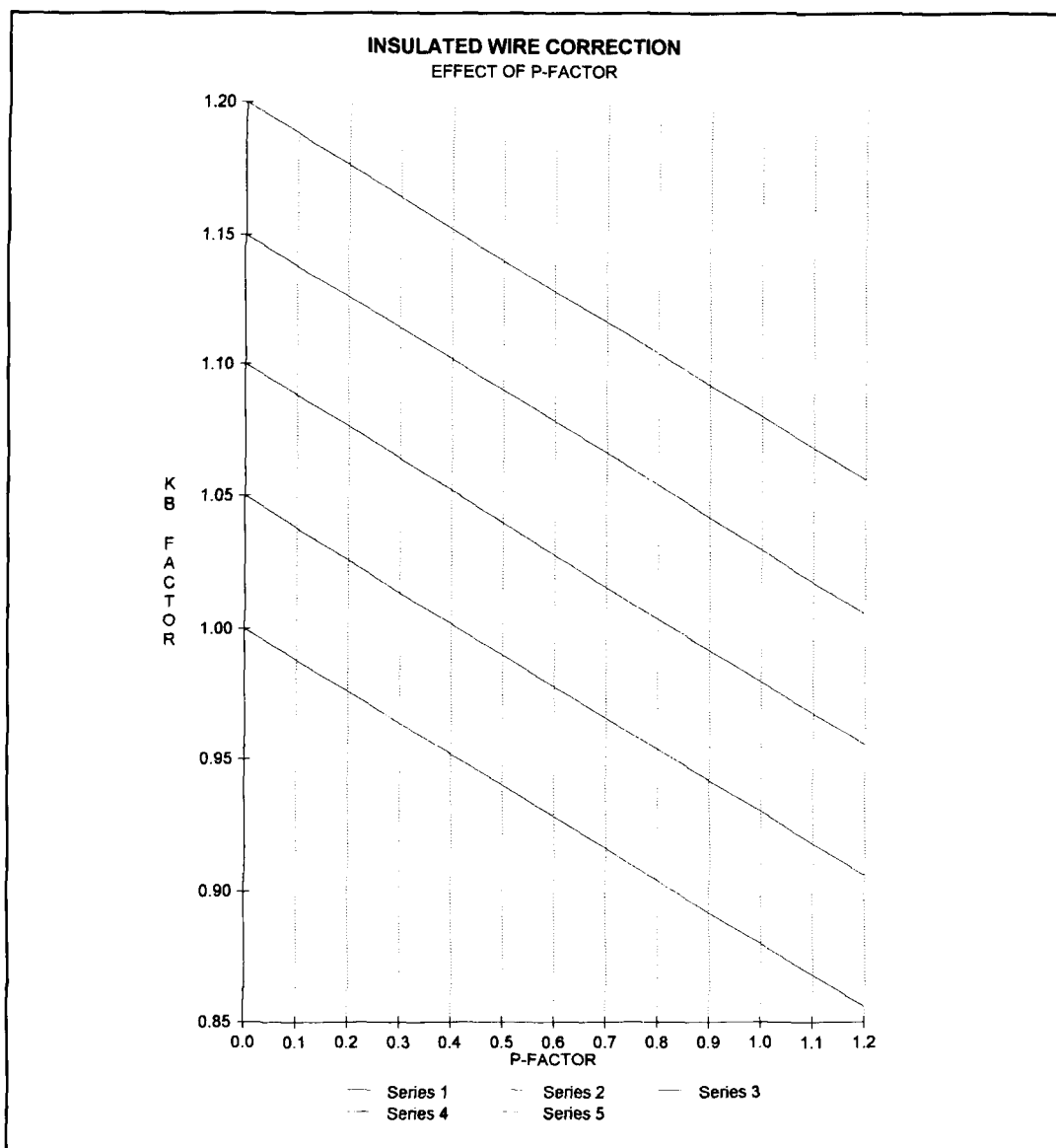
Figure 37. P-factor for insulation on antennas.

First, a parameter p is determined, a function of the insulation thickness and dielectric constant. The value for a specific insulation may be read from **Figure 37**. Next, the resonant frequency factor for the loop is determined, say from **Figure 13**. This value is used to enter **Figure 38** on the left axis, interpolating as necessary. Moving on the diagonal line (or parallel to it) to the value of p-factor provides the new resonant frequency factor by projection to the left axis. For example, if the p factor is 0.2 and the bare wire resonant factor is 1.10, the resonant factor with insulation is about 1.076. For a loop nominally resonant at 14.2 MHz, this amount of insulation moves the resonant frequency to about 13.89 MHz. The effect of this amount of insulation on forward gain will be

very small, but the point of maximum F/B will move by about the same amount.

A better way to account for insulation is to build a full-size loop. measuring the resonant frequency. Wire length is then reduced to get the desired resonant frequency.

Future articles

The next article in this series will be devoted to the two-element quad. Later ones will cover multi-element quads, design variations, and the effect of ground. ■

REFERENCES
1. R.P. Haviland. W4MB. "The Quad Antenna, Part 1: General Concepts,"
Ham Radio, May 1988, page 43.
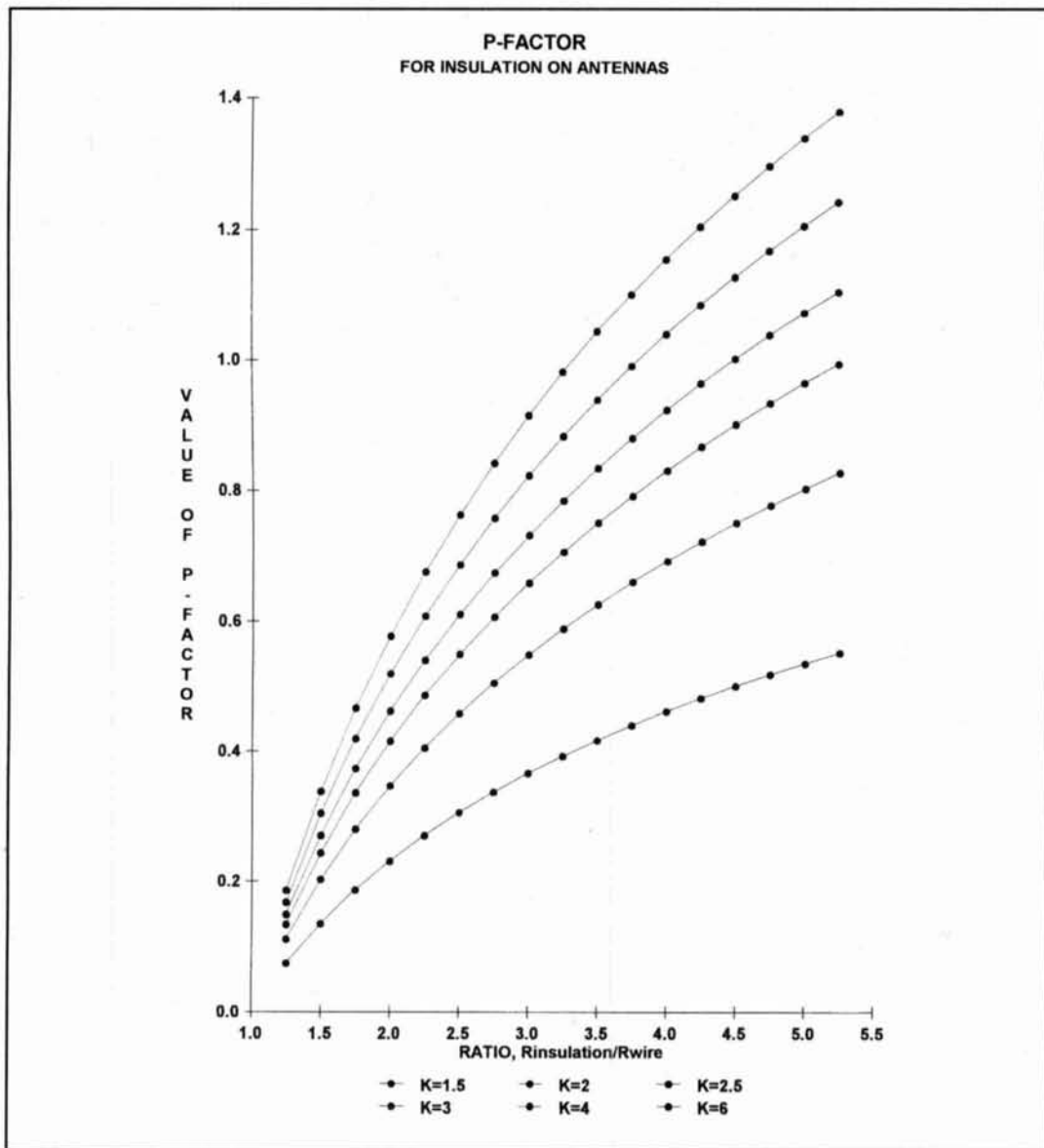
Figure 38. Insulated wire correction for quads.

2. Bob (R.P.) Haviland, W4MB, *The Quad Antenna*, CQ Communications, Hicksville, NY, 1993.

3. A.J. Julian, J.C. Logan, and J.W. Rockway, *MININEC: a Mini-numerical Electromagnetics Code*, Naval Ocean Systems Center TD 516, September 1982.

4. *NEEDS, the Numerical Electromagnetic Engineering Design System*, v2, June 1989. Available from The Applied Electromagnetics Society, R.W. Adler, Code 62AB, Naval Postgraduate School, Monterey, California 93943

5. *Computer Program for Thin-Wire Structures in a Homogenous Conducting Medium*, J.H. Richmond, Ohio State University, Columbus, Ohio, 1974.

6. R.P. Haviland, W4MB, "Ground Parameters for Antenna Analysis," *ARRL Antenna Compendium*, No. 5, 1996, American Radio Relay League, Newington, Connecticut.

7. R.P. Haviland, W4MB, "Insulated Antennas," *Communications Quarterly*, Winter 1993, page 75.

8. R.P. Haviland, W4MB, "Attaining Design Performance in Parasitic Arrays," *Applied Computational Electromagnetics Society Newsletter*, 13,1, March 1998.

9. Fouad Antoine Raphael, "Loop Yagi Array Antennas," Texas A&M Graduate Thesis, December 1979.

PRODUCT INFORMATION

New "El Toro" Synthesizers from Anritsu

Anritsu Company introduces the enhanced "El Toro" family of high-performance, ultra-low-noise microwave/millimeter-wave synthesized signal generators.

The new 68C and 69B "El Toro" product line consists of 42 models, including the 6937B synthesized generator that provides 0.1 Hz to 65 GHz frequency coverage in one instrument.

For more information, contact Anritsu Company at 1155 E. Collins Blvd., Richardson, Texas 75081 or visit their website at <www.anritsu.com>.

Rick Littlefield, *K1BQT*
109A McDaniel Shore Drive
Barrington, New Hampshire 03825

DISCPOLE ANTENNAS AND PROXIMITY TO GROUND

Investigating the effect of ground mounting

The "Discpole" antenna, which has been well documented in several publications,[1,2,3,4] is a capacitively loaded off-center fed (OCF) vertical dipole with a 50-ohm feedpoint at its base. Its performance has been described at mounting heights of 1/4-wavelength and above, over a frequency range of 14 to 450 MHz. Although off-center fed, this antenna is, by definition, a balanced radiator exhibiting symmetrical current distribution in a free-space environment.

Mounting instructions in the literature describing various versions of the design have all pointed out the need for adequate separation between the lower capacitive hat and ground in order to prevent loss of symmetry and detuning. At frequencies above 14 MHz, the requirement is easily met. Below 14 MHz, maintaining adequate ground separation becomes increasingly problematic.

In his recent article "Another Way to Look at Vertical Antennas," Rudy Severns, N6LF, described a 160-meter discpole-like project constructed at his home QTH.[5] This article piqued my interest, not only because it discussed an antenna that was similar to the discpole, but also because it specified near-ground-level mounting for the antenna's lower capacitive-loading structure. This was somewhat

troubling because, in previous tests, I've observed that discpoles mounted below 1/4-wave above ground level (AGL) exhibit the following qualities:

1. A rapid increase in VSWR with closer proximity to ground,
2. A downward shift in resonant frequency,
3. A skewing of the current distribution toward the base.

Based on these observations, I'd concluded that the discpole isn't particularly well suited for ground mounting. However, after reading Rudy's positive article, I decided to investigate further to see if I could generate some real-world data to confirm the antenna's viability at ground level. I was especially interested in its performance over lossy ground, that unpleasant condition most of us face with our backyard installations.

While computer modeling is often cited (and arguably overused) to predict antenna performance, it is well known that NEC-derived antenna programs may not always deliver accurate predictions for wires positioned in close proximity to lossy ground—a consideration Rudy's article thoughtfully pointed out. In the debate over elevated radials, W8JI and others have postulated that actual field-strength read-

Photo A. The test setup.

ings may represent a superior method for determining performance, especially when low portions of the antenna are operated near lossy media.[6] Because of this, I chose field strength readings as a yardstick for evaluating the discpole near ground.

Test setup and procedure

Building a discpole for 1850 kHz is a massive undertaking. Wire capacitive hats must be used instead of solid discs, tall vertical supports are needed, and a hefty patch of real-estate is required to make it happen! Lacking Rudy's fortuitous access to a stand of 150-foot Douglas fir trees on a ridge top, I elected to conduct my test on a patch of flat moist sandy soil at 146 MHz. This "experiment-in-miniature" offered considerably more control over the test environment. The ground surface I chose had less detuning effect than a similarly placed metal screen, but enough influence to impact VSWR and tuning—much as soil might in the real world at lower frequencies.

My test antenna was mounted on a movable wooden arm that also supported a coaxial choke balun and a length of feedline leading to small handheld antenna analyzer (see **Photo A**). I affixed this arm to a vertical length of plastic pipe using a U-bolt assembly that allowed rapid adjustments between 1/4-wave AGL and ground level. The feed system (balun plus patch cable) was pruned to one electrical wavelength to ensure congruence between its near and far

ends. Antenna VSWR at the 146-MHz test frequency measured a flat 1:1 at 1/4-wave AGL and in free space. The feedline's outer surface was, for all intents and purposes, non-radiating and isolated from the antenna.

The analyzer was used to measure VSWR and to generate a signal for field-strength measurements. Field-strength measurements were conducted from two monitoring locations: one at five wavelengths and the other at eight wavelengths from the test site at approximately 10-degrees inclination. I used two MFJ-224 signal analyzers, both outfitted with short vertical voltage-probe antennas adjusted for mid-scale meter deflection. The MFJ-224 provides 1-dB resolution over a 60-dB measurement range, and both units were checked against a HP-8640B prior to testing for calibration. I made multiple test runs, and I swung the arm supporting the antenna to different points about its insulated mounting post to help deflect unspecified variable and to ensure a wide range of data points.

Results

When the discpole was repositioned from 1/4-wave AGL to ground level, the observed field strength consistently dropped by 1 to 5 dB. Over a run of 15 trials, the mean decrease in performance was calculated at 2.6 dB. In addition to this performance decrease, VSWR typically increased from 1:1 at 1/4-wave AGL to between 1.5:1 and 2:1 at ground level

(although impedance magnitude stayed fairly constant). As predicted from earlier experiments, the antenna's resonant frequency shifted downward—a condition also noted by N6LF with his 1.8-MHz antenna.

In all fairness, data gathered at VHF may not be particularly valid for predicting HF antenna performance due to the tremendous differences in soil skin depths between the two frequency ranges. However, the fact remains that the data I collected portrayed a consistent tendency for performance to suffer when the discpole was positioned at ground level over lossy soil.

Discussion

The case for impaired discpole performance at ground level makes sense from a theoretical perspective. Like a dipole, the discpole is an inherently balanced radiator. When a balanced radiator is tipped on end and lowered voltage-loop-first toward lossy ground, the impedance of the lower half-element drops far more rapidly than the top half. This condition quickly destroys electrical symmetry, altering the radiator's pattern and take-off angle (TOA), VSWR, and resonant frequency. Furthermore, as the impedance differential widens with proximity, progressively more RF energy is drawn to the lossiest part of the system. If our eyes were capable of visualizing the shift that occurs in RF distribution as the antenna approaches ground, it wouldn't be a pretty sight.

On the other hand, the field-strength numbers also have a positive spin because they suggest that a carefully constructed discpole may yield close to 0-dBi performance when mounted at ground level (2.14 dBi - 2.6 dB loss). On 160 meters, this result isn't too shabby for a short vertical operating without benefit of an exten-

sive radial system. I'd proffer that decreasing the area of the bottom hat and increasing the area of the top may serve to reduce the impact of lossy ground to a minor extent, and to improve performance slightly.

Finally, after looking at this data, I don't believe I would install *any* inherently balanced off-center fed discpole or "elevated-feed"-type radiator with one end too close to lossy dirt (Cushcraft, Gap, MFJ-Ameritron, and Butternut verticals fall into this category). A ground-mounted 1/2-wave design *will* get you on the air with a re-spectable signal, as N6LF has so aptly demonstrated. But, all of these antennas will probably work better with some altitude under them.

Conclusion

A 15-trial field-strength experiment conducted at 146 MHz suggests that a discpole antenna loses 2.6 dB in low-angle performance when moved from 1/4-wavelength above ground level to ground level. The transferability of this data to 1.8 or 3.7 MHz is not known, and some have argued that no valid comparison can be made. However, at VHF, the trend toward degraded performance is both well defined and supported in theory. ∎

REFERENCES

1. Rick Littlefield, K1BQT, "The 2-Meter Discpole Antenna," *Communications Quarterly*, Summer 1996, pages 77–81.
2. Rick Littlefield, "Build the Six-Meter Discpole Antenna," *CQ VHF*, March 1997, pages 44–50.
3. Rick Littlefield, "The Discpole Antenna," *RF Design*, May 1997.
4. Rick Littlefield, "Build a 20-Meter DX-pole Antenna," *Communications Quarterly*, Spring 1997, pages 98–102.
5. Rudy Severns, N6LF, "Another Way to Look at Vertical Antennas," *QEX*, March/April 1999, pages 3–8.
6. Tom Rauch, W8JI, "Comments on Elevated Radials," Technical Conversations, *Communications Quarterly*, Summer 1998, pages 6, 7, 106, and 107.

PRODUCT INFORMATION

Hi-Res Satellite Images for Windows

Timestep, manufacturers of equipment for receiving live images from weather satellites, has launched their latest Windows product. *HRPT for Windows* is a new system for resolving the high-resolution (1.1 km) digital images for NOAA 12,14, and 15 and is completely Windows 95, 98, and NT compatible.

The program is highly automated, including automatic scheduling to eliminate the need to predict start times, automatic channel switching for continuous unattended reception of the three satellites, and automatic Internet connection for element updates.

The software saves all five-band 10-bit data and provides real-time 30-bit MultiSpectral

color reception and manipulation; preset MultiSpectral band allocations give faster color; other features combine to produce clear images showing even small towns and rivers.

HRPT (high-resolution picture transmission) is digital data from the NOAA polar orbiting satellites. Timestep also produces hardware and software for receiving standard analog APT (automatic picture transmission) from all polar orbiting satellites, as well as for Meteosat, GOES, and other geo-stationary satellites.

For more information about Timestep's *HRPT for Windows* and their other products, visit their website at <http://www.Time-step.com> or send an e-mail to <sales@Time-step.com>.

Rick Littlefield, *K1BQT*
109A McDaniel Shore Drive
Barrington, New Hampshire 03825

QUARTERLY DEVICES

The Analog Devices AD8307 92-dB logarithmic amplifier

The 90-dB dynamic range of a good spectrum analyzer sure beats a simple diode-driven meter for sniffing out RF. Yet, this level of sophistication is normally unavailable for simple in-line monitors and hand-held testers. The AD8307, a new wide-band detector and logarithmic amplifier IC from Analog Devices, may change that. With microvolt sensitivity, 92-dB dynamic range, 1 to 500 MHz frequency coverage, and a low price tag, the AD8307 packs a lot of instrumentation solutions into a tiny 8-pin package!

History

Most experimenters are familiar with the RSSI (receiver signal-strength indicator) feature built into FM receiver ICs. This function delivers logarithmic signal-strength data that may be used to drive an S-meter, activate a channel-in-use indicator, control a multi-level squelch, or notify scanning logic of channel activity. Cell phones encode RSSI levels and return them to the cell site to help it establish transmitter power and hand-off times.

While useful for FM and cell phone communication, log-amp RSSI levels are also adaptable to many other purposes. That's because they provide a *logarithmic* rendering of incoming signal strength that allows linear-scale meters and oscilloscope grids to be calibrated directly in dB or dBm. This property is great for any application where RF energy must be measured over a wide dynamic range and expressed in dB. For some time, experimenters and designers have misappropriated FM-IF chips like the NE604 precisely for this function, using 18 or 24-pin multi-function devices to harness the RF log-amp inside.
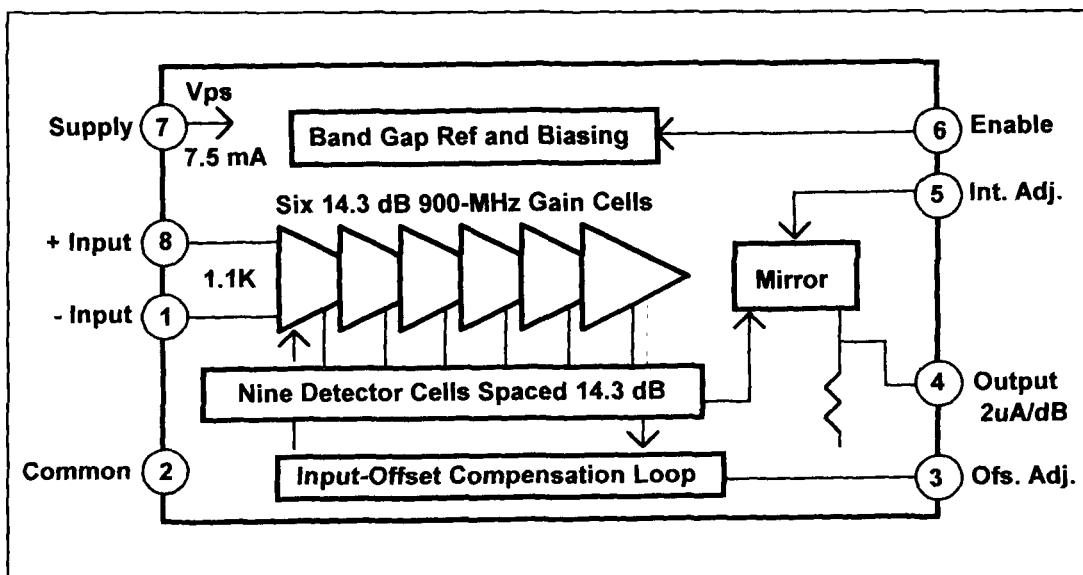
Figure 1. Functional block diagram for the AD8307.
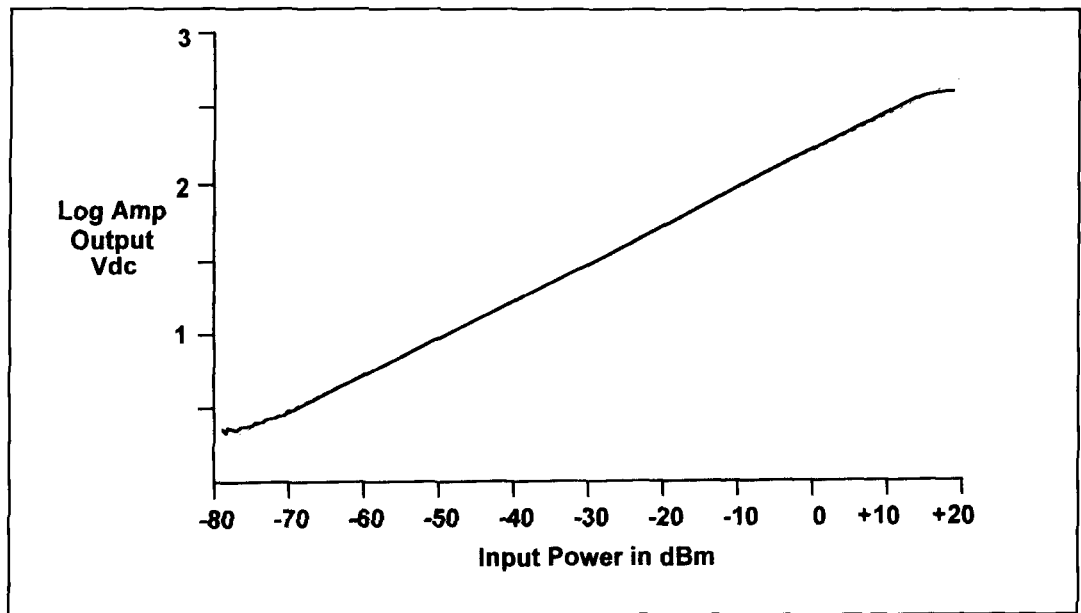
Figure 2. The AD8307 remains linear to within 1 dB over most of its 93-dB dynamic range.

While the AD8307 provides RF designers with the same general log-amp function found in those devices, there are several pluses that make it more of an instrumentation-grade rather than a consumer-grade item. For one thing, it's packaged in a much smaller dedicated 8-pin chip that features lower power consumption. Beyond that, it offers better dynamic range, greater frequency span, plus the ability to adjust internal offset and intercept parameters. In other words, it's the right tool for the job rather than an adaptation of a consumer part!

Product description

For a look at what's inside the AD8307, refer to **Figure 1**. The 1.15-k differential input port provides a usable detection span of 1 to 500 MHz. However, frequency coverage may be extended upward to 900 MHz or downward to DC with some sacrifice in dynamic range. You may also add filtering at the front-end to restrict the passband and eliminate undesired pickup. The AD8307's useful amplitude-measurement range is 92 dB, with a threshold sensitivity of - 75 dBm (extendible to -90 dBm with external impedance matching). For especially demanding applications, the addition of a AD603 front-end increases dynamic rangeto 120 dB!

To realize its impressive dynamic range, the AD8307 uses a process called *successive detection*. The theory behind successive detection is quite complex and is covered at length in the device's comprehensive 20-page data and application sheet. However, to quickly summarize and grossly understate what happens, sig-

nals are fed into a chain of six cascaded 14.3-dB amplifier/limiter cells. Left to run wide-open, these cells develop a lot of total gain—in excess of 85 dB. However, developing raw gain isn't the primary mission or desired function of the cascaded log cells. Instead, their job is to work together to provide *controlled gain compression* of incoming signals over a wide range of possible input levels.

To implement this balancing act, nine detector cells constantly monitor signal levels in each cell and compile a summary signal for the device's feedback network. At the far end of the device, a mirror circuit uses this information to generate a DC output level that accurately depicts how much gain compression is needed at any point in time to "keep the lid" on incoming signals. In addition to the detector network, the device also has an off-set compensation loop to sense DC levels and keep RF signals between the rails in each cell. The cells are all DC-coupled, so each successive amplifier must be compensated.

Below 100 MHz, accuracy of the AD8307 is typically within 1 dB over 88 dB of its range. The RSSI output signal normally rests at 0.250 mV (-74 dBm) and increases at the rate of 25 mV per dB as signals rise out of the noise (see **Figure 2**). In practical terms, this output level is adequate to drive a zero-compensated 50-µA meter and most forms of electronic sensing directly, but not quite stiff enough to drive a low-cost 200-µA meter movement without a buffer.

Like many new devices entering the "low-voltage portable" market, the AD8307's power consumption is next-to-nothing! Usable Vps operating range is 2.7 to 5.5 volts DC with a

```
IN-  □    ●    □  IN+
Com  □  AD8307  □  Vps
Ofs  □   Top    □  Enb
           View
Out  □         □  Int
```
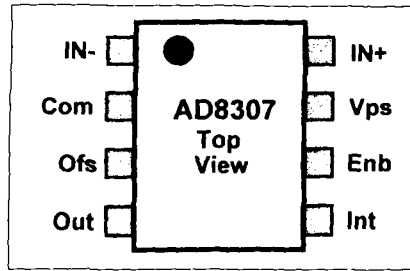
AD8307—SPECIFICATIONS (V_s = +5 V, T_A = 25°C, R_L ≥ 1 MΩ, unless otherwise noted)

Parameter	Conditions	Min	Typ	Max	Units
GENERAL CHARACTERISTICS					
Input Range (±1 dB Error)	Expressed in dBm re 50 Ω	-72		16	dBm
Logarithmic Conformance	f ≤ 100 MHz, Central 80 dB		±0.3	±1	dB
	f = 500 MHz, Central 75 dB		±0.5		dB
Logarithmic Slope	Unadjusted[1]	23	25	27	mV/dB
vs. Temperature		23		27	mV/dB
Logarithmic Intercept	Sine Amplitude; Unadjusted[2]		20		μV
	Equivalent Sine Power in 50 Ω	-87	-84	-77	dBm
vs. Temperature		-88		-76	dBm
Input Noise Spectral Density	Inputs Shorted		1.5		nV/√Hz
Operating Noise Floor	R_SOURCE = 50 Ω/2		-78		dBm
Output Resistance	Pin 4 to Ground	10	12.5	15	kΩ
Internal Load Capacitance			3.5		pF
Response Time	Small Signal, 10%-90%,		400		ns
	0 mV–100 mV, C_L = 2 pF				
	Large Signal, 10%-90%,		500		ns
	0 V–2.4 V, C_L = 2 pF				
Upper Usable Frequency[3]			500		MHz
Lower Usable Frequency	Input AC-Coupled		10		Hz
AMPLIFIER CELL CHARACTERISTICS					
Cell Bandwidth	-3 dB		900		MHz
Cell Gain			14.3		dB
INPUT CHARACTERISTICS					
DC Common-Mode Voltage	Inputs AC-Coupled		3.2		V
Common-Mode Range	Either Input (Small Signal)	-0.3	1.6	V_S - 1	V
DC Input Offset Voltage[4]	R_SOURCE ≤ 50 Ω		50	500	μV
	Drift		0.8		μV/°C
Incremental Input Resistance	Differential		1.1		kΩ
Input Capacitance	Either Pin to Ground		1.4		pF
Bias Current	Either Input		10	25	μA
POWER INTERFACES					
Supply Voltage		2.7		5.5	V
Supply Current	V_ENB ≥ 2 V		8	10	mA
Disabled	V_ENB ≤ 1 V		150	750	μA

NOTES
[1]This may be adjusted downward by adding a shunt resistor from the Output to Ground. A 50 kΩ resistor will reduce the nominal slope to 20 mV/dB.
[2]This may be adjusted in either direction by a voltage applied to Pin 5, with a scale factor of 8 dB/V.
[3]See Application on 900 MHz operation.
[4]Normally nulled automatically by internal offset correction loop. May be manually nulled by a voltage applied between Pin 3 and Ground; see APPLICATIONS.
Specifications subject to change without notice.

Figure 3. AD8307 top view.

nominal current drain of only 7.5 mA. In sleep mode, the device nods off at 150 μA. Power-up time is 100 nS. The AD8307 is quite insensitive to changes in supply voltage and temperature, making it a natural for battery-operated in-the-field applications. Complete data is provided in **Figure 3**.

Applications

Potential applications for the AD8307 run a wide gamut. Many are obvious, such as RF power meters, spectrum analyzers, and service monitors. Others include true-dBm level meters for RF and audio generators, EMR and electric field interference detectors, and radar and sonar signal processors. The AD8307 should prove useful anywhere it's desirable to compress wide-ranging signal levels into a decibel-based display.

One feature I particularly like about the AD8307 is its self-contained simplicity. You won't need to add a lot of external parts to make it play. An illustration of this appears on
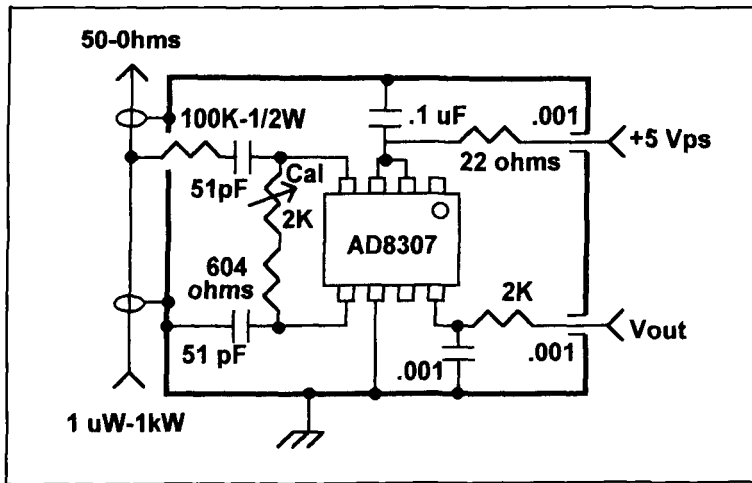
Figure 4. Circuit diagram illustrates use of the AD8307 in a simple wide-range in-line wattmeter. Power range is 1 µW to 1 kW without range switching.

page 17 of the application note, reproduced here in **Figure 4**. This bare-bones RF-power in-line meter circuit spans a wide frequency range and also covers a 90-dB power range, taking you from sub-milliwatt levels up through 1 kW without need for a range switch. A resistive sample tap attenuates the through-line signal, bringing it down to the device's operating range. Shielded packaging prevents stray RF pick-up, and feed-through caps isolate the power and signal lines. The unprocessed RSSI output signal may be transmitted by cable to a remote meter or DVM-type display located some distance away from the sensor. For raw simplicity, this circuit is pretty hard to beat! Simplicity is always welcomed by product designers, because a low parts count usually means more profit and greater reliability.

For an equally simple application, see the circuit illustrated in **Figure 5**. This is essentially an amplified RF-field-strength meter. Potential
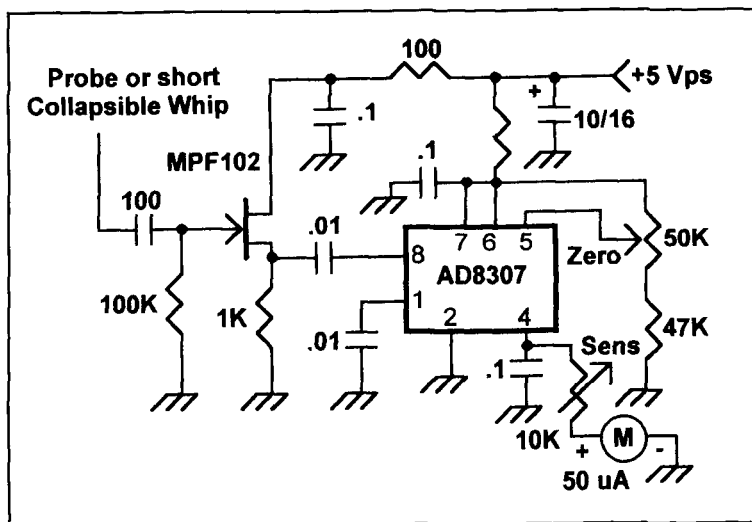
applications include a transmitter tester, antenna E-field indicator, cable-leakage detector, counter-surveillance monitor, RF-noise locator, and oscillator checker. It responds to a much wider amplitude range than a simple diode-driven meter, so it has more uses. A simple FET-follower circuit matches the voltage probe's relatively high impedance into the 1.1-k impedance of the log-amp. Because of the device's wide frequency and amplitude range, a shielded enclosure is strongly recommended. Power may be supplied by a 9-volt flat-pack battery conditioned through a monolithic regulator IC or 5.1 zener.

Generally speaking, it's important to use good VHF layout and construction practices— and also to provide both high and low-frequency bypassing—when working with this device. Also, because it amplifies efficiently over a DC-to-daylight frequency span, you'll need to pay attention to what you allow to pass into the front-end (60-Hz hum, for example). However, these simple precautions shouldn't suggest that the AD8307 is squirrely or difficult to use. On the contrary, while breadboarding a simple "get-aquatinted project" with DIP packaging, I found the device both stable and pleasantly forgiving! The AD8307 is available in 8-pin DIP and SOIC packaging, making it a convenient choice for both experimental bench projects and mass-produced SMD designs.

AD8307 technical information and pricing

To download a comprehensive data sheet for the AD8307, visit Analog Device's website at <http://www.analog.com/logamps>. Estimated pricing is $7.10 in 100-lot and $6.00 in 1,000-lot quantities. If additional information or sampling is needed, you may write to Analog Devices at One Technology Way, P.O. Box 9106, Norwood, Massachusetts 02062-9106 or call them toll-free at 1-800-ANALOGD (1-800-262-5643).

Conclusion

The AD8307 is a low-cost dedicated multi-stage log-amp for RSSI and RF power measurement applications. It features a 92-dB dynamic range, 1-dB linearity, and -75 dBm sensitivity over a DC-to-500 MHz bandwidth. Available in 8-pin PDIP or SOIC packaging, the AD8307 operates from a single low-voltage/low current supply with excellent stability over temperature and voltage fluctuations. This device appears to be both "forgiving" and easy to implement, and it should find wide acceptance for a variety of low-cost RF-instrumentation applications. ∎

Figure 5. Simple RF field-strength meter uses active voltage-probe antenna and AD8307 RF log amp to measure an unusually wide range of signals.

Walter J. Schulz, *K3OQF/VQ9TD*
P.O. Box 4054
Jim Thorpe, Pennsylvania 18229

THE LAST 100 YEARS IN COMMUNICATION

Part 2: The transition

While the North Atlantic wireless traf-
fic between America and the
European continent remained on the
low frequencies, noise and static interference
made communications on these wavelengths
difficult. It was discovered early on that north-
ern latitudes yielded better radio conditions—
a reduction of noise and static interference—
because the great circle distance was shorter.
However, that wasn't the total answer to the
reception problem. A better receiving antenna
system was needed.

Enter Harold Beverage

As a young boy in Maine, Harold H.
Beverage built a homebrew wireless receiver.
With it, he heard the rescue ship *S.S.
Carpathia*'s Morse code signals as the she sped
toward the sinking *Titanic* in the north Atlantic
Ocean. After graduating from the University of
Maine in 1915, Beverage landed a job with the
General Electric Company and was eventually
transferred to its radio laboratory headed by Dr.
E.F. Alexanderson.

During World War I, Alexanderson was
assigned the job of improving reception of the
Barrage Receiver. Testing was done at New
Brunswick, New Jersey, and at the Naval Radio
Station NBD at Otter Cliffs near Bar Harbor,
Maine. The Barrage Receiver used long anten-
na wires laid out along the ground. Dr.
Beverage noticed that the antenna's reception
pattern seemed to be directive. This antenna
system probably provided the idea for
Beverage's invention of the wave antenna.

**Photo A. The first host of Radio Station WEAF, also the announcer for the AT&T
radio station, was Miss Helen Hann, a member of the Long Lines Department.
This picture was taken fall of 1922. (Courtesy of AT&T)**

Beverage continued to experiment with long-
wire antennas stretching out over approximate-
ly six miles on Long Island, New York. At that
time he worked for the Radio Corporation of
America's (RCA) communication division, and
was seeking a way to reduce noise and static
while receiving longwave signals. On Long
Island, Dr. Beverage and his colleague Philip
S. Carter discovered that signals from Europe
increased on a particular long-wire antenna
strung along a road running northeast to south-
west. Carter did a complete mathematical
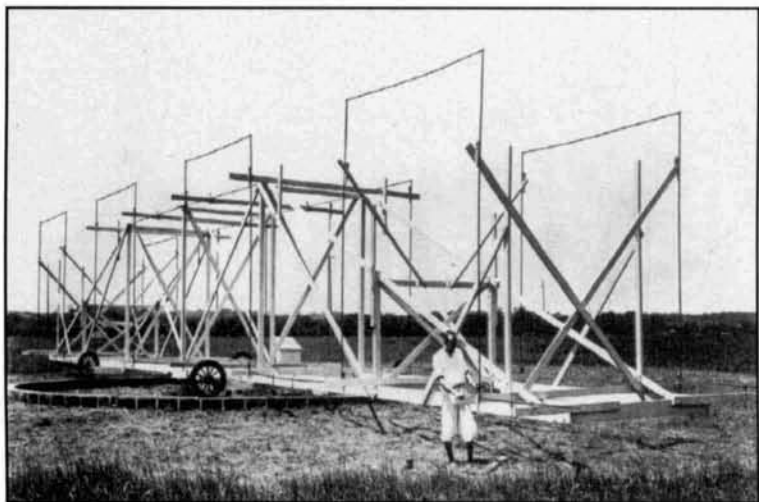
Photo B. The late Karl G. Jansky of Bell Telephone Laboratories, with the rotation Bruce Antenna he used to discover radiowaves coming from outer space. Mr. Jansky's investigations during the 1930s lead to this discovery and gave the world a new science—radio astronomy. (Courtesy of Bell Telephone Laboratories)

analysis and confirmed their conclusions about the wire's directive characteristics.

The Beverage or Wave antenna, as we call it today, was patented on June 7, 1921. This antenna has become a world standard for receiving. It is often used on low frequencies and even on radio amateur bands 1.8 and 3.5 MHz due to its pronounced directivity and static reduction capabilities.

The vacuum tube

Lee De Forest, from Council Bluffs, Iowa, was a graduate of the Sheffield Scientific School at Yale. He improved upon Sir Ambrose Fleming's two-element valve (diode) by supplying a third element, called a grid, between the diode cathode and plate. The new device was called the Audion. Unfortunately, the tube didn't have a long life span and usually burned out after only short-term use.

Photo C. The Late Mr. Edmund Laport on left and the author on the right at Mr. Laport's farm in Ringoes, New Jersey.

It wasn't until Harold Arnold, working for the Western Electric Company in 1913, altered the internal vacuum of the Audion (triode tube) that its failure rate changed significantly for the better. Arnold theorized that the Audion lacked a good vacuum. Through new vacuum pumping techniques, which were his specialty, Arnold improved the tube life with a high vacuum. This development greatly increased tube performance.

American Telephone and Telegraph (AT&T) bought the patent rights to De Forest's Audion, hoping to benefit from the device's amplification capability. AT&T wanted to amplify coast-to-coast telephone calls and found the device could also generate radio waves via oscillation. Edwin H. Armstrong's black box device of 1912 made this possible. Armstrong had been able to take the vacuum tube output and return it to the input circuit, producing positive feedback—a process then called regeneration. Adjusting the amount of regeneration made the tube amplify or oscillate. This single event made it possible to span oceans without alternators or arc converters.

In 1907, DeForest had shown that voice signals could be transmitted over the wireless. However, the improved triode vacuum tube that made the technology truly possible was introduced on April 4, 1915. Bell Telephone Engineers were able to transmit a voice 250 miles between Montauk Point, Long Island and Wilmington, Delaware. Then, on October 21, 1915, in the first transatlantic voice transmission, "Good night Shreeve" was heard by H.E. Shreeve who was positioned to listen at the Eiffel Tower in Paris. The originating signal was transmitted from a large antenna at the Naval Radio Station NSS at Arlington, Virginia, by B.B. Webb, R.A. Heising, and H.W. Everitt.

The actual transmitter for this historic event contained hundreds of 15-watt vacuum tubes. These tubes were connected together to produce adequate transmitting power for the signal to reach Europe. During the same year, radiotelephony signals were received by other Bell Telephone experimental sites located at various U.S. Naval installations in Honolulu, Hawaii, the Panama Canal Zone, Mare Island (San Francisco), and San Diego, California. These locations later were to become U.S. Naval Communication Stations: Honolulu became NPM, the Panama Canal Zone was known as NBA, Mare Island's callsign was NPG, and San Diego Naval Radio Station was NPL.

Before the first transatlantic radiotelephony communication between Naval Radio Arlington and Paris, France, the Arlington station was used primarily for wireless continuous wave (CW) message traffic. The radio station was commissioned for active naval service on February 13, 1913. It's callsign, NSS, was rec-

ognized around the world. The Naval Radio Station was located in Arlington, Virginia, across the Potomac River from Washington, D.C. Its high-power arcs and spark transmitters could always maintain communications with ships at sea, passing message traffic from ship to shore and vice-versa.
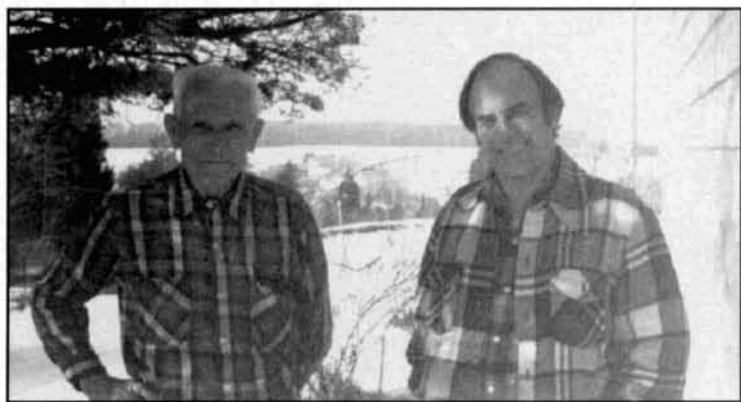
The Naval Radio Station's three main transmitting towers were built by the Baltimore Bridge Company. The towers made the station recognizable from far away because they dominated the horizon. One tower stood at least 600 feet tall. These heights gave the station a transmitting range of at least 3,000 miles. At that time, the towers were series-fed and used gigantic marble slabs as insulators at the base.

At the beginning of the wireless era, Radio NSS used a Fessenden-type 100-kW spark transmitter operating at the 2100-meter wavelength (143 kHz). A gigantic manually operated antenna switch was used for the transmitter and receiver changeover. As part of its equipment, the station also had Federal Electric Arc Converters and 5-kW transmitters. After the first transatlantic telephone call, the 100-kW spark and 5-kW electric arc transmitters were supplanted by new vacuum-tube transmitters in 1925. When 1928 arrived, Radio NSS was using 10-kW transmitters operating on the short wavelengths. The station was finally shut down for good in July 1956, making way for a new airplane runway.

During 1922, there were only 50 radio broadcast stations operating in the 300- to 400-meter band. In 1923, the number increased to 500 radio stations operating in this limited band. Two years later, radio stations KDKA in Pittsburgh, Pennsylvania, and WGY in Schenectady, New York, began to experiment with higher transmitter power, increasing their respective coverage areas.

The crowded and chaotic broadcast band conditions finally forced President Calvin Coolidge to ask the U.S. Congress to provide some relief. Congress passed the 1927 Radio Act, which created the Federal Radio Commission (FRC) to act as a regulatory body. The Commission had the authority to establish station classes, clear channel, regional channel, and local channel frequencies. It was also granted power to further regulate the new broadcast industry. The FRC assigned operating frequencies, and coordinated protection of stations from other stations' interference. In 1937, the Federal Communications Commission (FCC) was created to take over the Federal Radio Commission's responsibilities.

The Germans take an interest

In 1924, across the Atlantic, the German government was busy realizing the value of short-
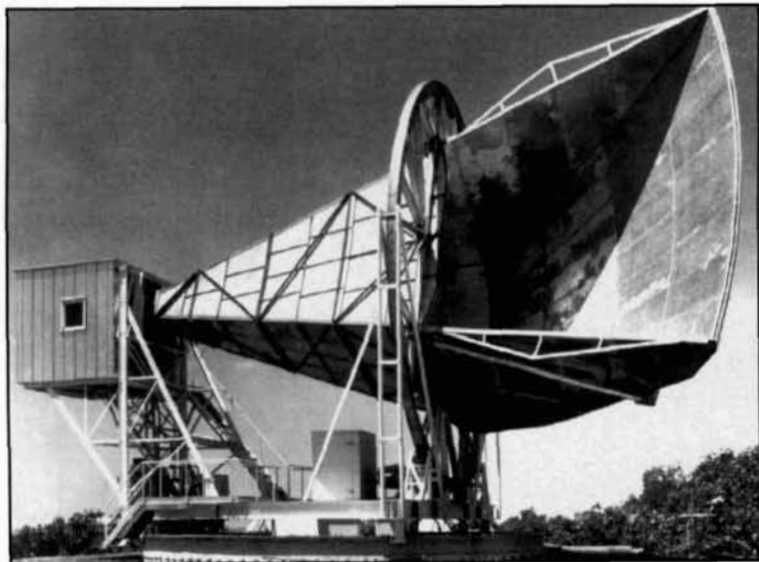
Photo D. Horn antenna used at Holmdel. It was a key element in Echo and Telstar Satellite experiments that later permitted scientists to discover background radiation of the universe. (Courtesy of Bell Telephone Laboratories)
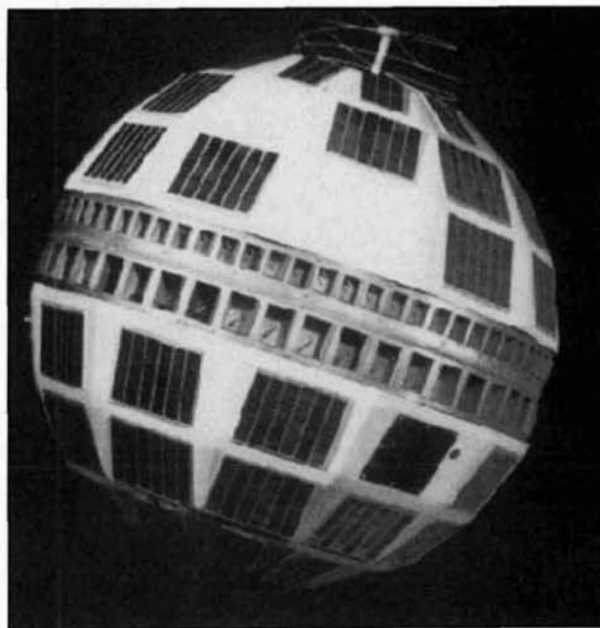
Photo E. Telstar Communication Satellite. (Courtesy of NASA)

wave communications and broadcasting. Argentina and Germany made their first short-wave contact on 4 MHz, although high-frequency usage didn't begin there. It's evident from the early wireless days that the Germans followed wireless developments with keen interest. In fact, some of their leading scientists were quite involved and were noted for their professional accomplishments in the wireless field—yet they were slow to change from long to short wavelengths.

The American Radio Relay League sent Paul Godley, 2XE, to Scotland to prove that transoceanic communications were possible
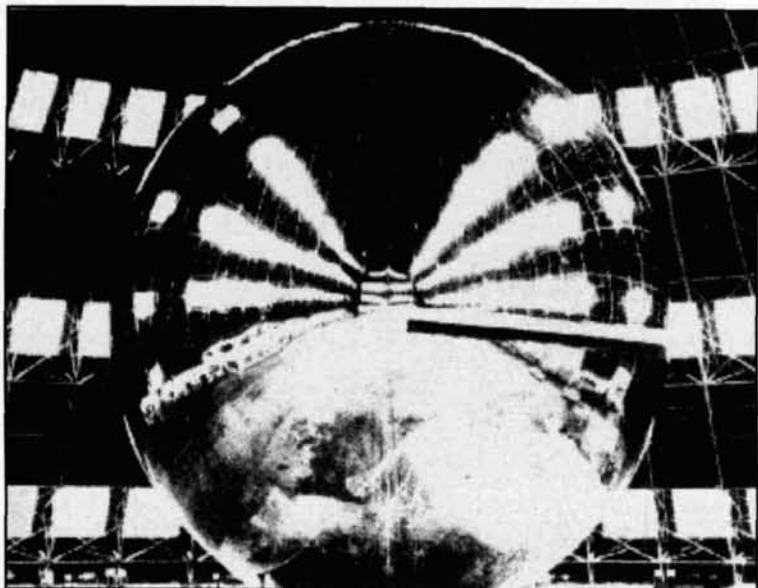
Photo F. Photograph of the Echo Balloon Satellite used to reflect radiowaves back to Earth for communications. (Courtesy of NASA)

using shorter wavelengths. Germany took notice and began its own development program using short waves. At the time, Mr. Godley was America's best. Besides being a ham radio operator, he also was a member of the Institute of Radio Engineers. Godley stationed himself on the Scottish moors and set up one of Dr. Beverage's wave antennas. With it, he was able to hear American amateur radio stations on the short wavelengths. For example, he heard 1BCG, where Edwin G. Armstrong was one of the radio operators. Armstrong had incorporated some of his own ideas into the radio equipment, which helped make the exchange with Godley successful. After taking note of these communications between America and Europe, the German Post Office became interested in short wavelengths.

An engineer at Telefunken built an 800-watt transmitter for operation on 40 and 90 meters. This marked the start of the first German Telegraphic shortwave service to South America on July 18, 1924. Between 1924 and 1928, the Reich Post, like the private company Telefunken, recognized the value of shorter wavelengths and began to develop communications facilities that used these wavelengths. At Nauen, a 20-kW transmitter was built using an antenna array with 48 dipoles.

German broadcasting was under the authority of the Deutsches Reich Post, the German National Post Office, and daily shortwave broadcasts originated from Nauen Radio and were sent to the U.S., South America, and the Dutch East Indies. Directional antenna systems were first used to beam radio programs to North America in 1932. Use of an antenna array improved signal reception quality. As

1943 approached, the Germans had shortwave radio stations at Zeesen, Oebisfelde, Munich-Ismaning, and Elmhorn. They were using transmitter output powers ranging from 50 to 120 kW feeding Rhombic and Curtain arrays.

Meanwhile, across the Atlantic

The Germans weren't the only ones who were beginning to realize the value of broadcasting programs and music via the radio. On August 10, 1922, a new radio station with the callsign WOO went on the air from Philadelphia's John Wanamaker Department Store. One of the programs presented was a March 15, 1923 concert—the first successful attempt at transmitting organ music. Marcel Dupre, an organist from the Notre Dame Cathedral, played the famous John Wanamaker store organ and was heard in Paris. Station WOO also transmitted President Calvin Coolidge's funeral address for President Warren G. Harding, who died in San Francisco on August 2, 1923. This WOO presidential broadcast was heard in England. However, the real developments in radio broadcasting were taking place at WEAF in New York City.

WEAF was the forerunner of today's radio networks. This experimental station was operated by AT&T (**Photo A**). It was the testing ground for radio broadcasts that transmitted various types of programming to public audiences. The public's favorable reaction was responsible for broadcasting's emergence as a new industry. RCA sold radio receivers, and its growth was meteoric, because the radio became a common American household item. However, it was the telephone company's unique talent for promotional techniques that made radio broadcasting commercially viable. AT&T had the ability to connect radio stations in distant cities via telephone lines; this enabled simultaneous broadcasts which, in essence, formed a network that made the technology profitable.

AT&T's techniques were so successful that other stations not included in their network deemed the phone company's practices unfair. This led to the eventual acquisition of WEAF by RCA and the formation of its own broadcast network under the banner of a new company, the National Broadcasting Company, or NBC, which employed the experienced WEAF staff.

Two offshoots of the WEAF experiment were the technologies of late modulation and oscillator circuits. These special circuits were developed by Mr. Heising and were called "constant-current modulation," referring to the "plate modulation circuitry." Colpitts and Hartley had developed the tube-oscillator circuits. All these special circuits were developed at the Bell Telephone Laboratories.

With the WEAF experiment completed, Bell Telephone Laboratories continued its research into radiotelephone transmission across the North Atlantic to Europe, as well as with ships at sea. In 1916, Bell engineers demonstrated successfully that radiotelephony could be used aboard ships. While sitting in his Washington, D.C. office, The Secretary of the Navy was able to talk with the *U.S.S. New Hampshire*'s captain. This two-way radiotelephone transmission was conducted between the nation's capitol and the ship, which was sailing off the Chesapeake Capes. The demonstration proved that radiotelephony could augment CW techniques and made it possible for the general public to have direct two-way voice conversations with people in European and aboard ships.

Bell Telephone Laboratories continued to investigate radiotelephone transmissions with their experimental radio station in Deal Beach, New Jersey, callsign 2XJ. The station invited wireless experimenters (radio amateurs) to listen for their test transmissions. Originally, it transmitted voice and music using Western Electric vacuum tubes. In the spring of 1920, transmitter power was increased to 1 kW at the shorter wavelengths, but the question remained as to what type of modulation was needed.

Physicist John R. Carson was employed by AT&T to solve a modulation problem. He came up with a mathematical analysis of amplitude modulation. From his analysis, basic principles were derived for transmitting single sideband voice modulation. Carson's contribution to modulation, along with the experiments carried out by E.L. Nelson and A. Oswald at 2XJ, made possible the first two-way voice commercial transmission from New York City to London, England on January 7, 1927.

John Wanamaker's Philadelphia store gave up radio station WOO in 1927, and AT&T acquired the callsign for its ship-to-shore radio telephone service. The company established the high seas radio telephone station at Deal Beach along with its associated reception station at Forked River, New Jersey. Later, the transmitter site was moved to a salt marsh just south of Tom's River at Ocean Gate, New Jersey. The receiver site was later moved yet again and re-established at Manahawkin, New Jersey. By 1934, there were many ships in the Atlantic Ocean using the radio telephone service established at this radio telephony station. The station was, and is still, the gateway to the U.S. and the world telephone network on the high seas.

This same station, along with its other ship-to-shore responsibilities, maintained point-to-point radio telephone communications circuits with 52 foreign countries. What made the station so important was that the Bell Telephone system continued its research to improve the quality of radio circuit transmissions and implemented the results at WOO. These continuing efforts led to the development of new antenna systems and to advancements in the study of propagation phenomena via stations whose transmissions spanned the North Atlantic circuits.

Bell Telephone Laboratories established a base of operations at Holmdel, New Jersey. This open farmland with rich soil served as a site for the development of the new antenna systems. Here many notable names in antenna development came together, initiating a new era with remarkable results in antenna engineering.

Directivity

As early as 1899, a proposal was made to use two vertical antennas physically separated at some fraction of a wavelength. One of the vertical antennas would be excited with a fraction of a given wavelength at a time, a bit later than the first vertical; this relationship would yield a directional radiation pattern as envisioned by S.G. Brown.

We've seen how Harold Beverage employed the traveling wave antenna to achieve a better quality signal using the directive properties of an antenna system. A wave-antenna array was constructed in Maine by AT&T to make use of these directive properties. It required large areas, but with the advent of shorter wavelengths, something else was needed to obtain an increased signal-to-noise ratio.

Around 1925, Bell Lab engineer Harold T. Friis began experimenting with a new means of obtaining directivity in the azimuthal plane. His goal was to reduce atmospheric interference by using two loop antennas separated by 1/12 of a wavelength. Friis believed that, with proper electrical phasing between the antennas, the system directivity would be increased. His method reduced static noise, and the antenna system used less land area than the Beverage antenna array.

At the same time Karl Jansky, also of Bell Labs, was looking for the source of static interference that plagued long-haul radio communication circuits. He built a rotatable Bruce antenna array (**Photo B**). Rotation was accomplished using an old model-T Ford engine placed in the nearby Holmdel meadow. Jansky was able to identify three different sources of static interference that plagued the North Atlantic transmission circuits. They were static crashes from nearby thunderstorms, static from very distant thunderstorms, and a steady hiss that seemed to have its origin in the cosmos. In 1933, Jansky wrote a paper entitled "Electrical Disturbances Apparently of Extraterrestrial Origin." It was based on established data col-
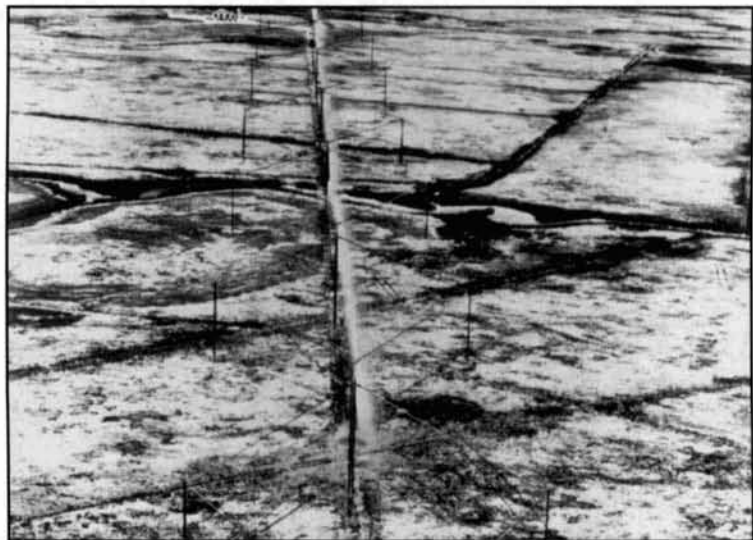
Photo G. Radio Telephony Station WOO used a rhombic antenna for reception, but, at a later date, a MUSA array was installed at Barnegut Bay, New Jersey. The system provided a link between England and the U.S. and employed 16 rhombics in a straight line directed at England on a great circle bearing. (Courtesy of Bell Telephone Laboratories)

lected during one year's research into the phenomena. The final conclusion was that the observed hissing noise came from the center of the Milky Way Galaxy.

In 1927, Edmond Bruce began to examine the wire antenna properties for radio communications. He found that tilting the wire antenna at certain levels increased signal voltage at the receiver input. As a result, the tilted-wire antenna evolved into the inverted-vee antenna (not to be confused with the inverted-vee dipole). This invention led to the diamond antenna used for shortwave receiving and transmitting.

Bruce, along with A. C. Beck and L. R. Lowry, further refined the engineering and design of the diamond antenna. Bruce's associate, Alfred C. Beck, had joined Bell Labs in 1928. Beck held 12 different patents for his various inventions in microwave transmission techniques and radar antenna development. His research led to the development of long-haul transmission circuits, microwave relay, and shortwave antenna systems.

Lowry had an interesting career with Bell Labs. He joined the Labs in 1927, and, in his early years with the company, helped develop the rhombic antenna system, along with Bruce and Beck. The rhombic antenna became a very popular system for commercial ship-to-shore stations as well as for Naval Communications. These antennas are still used around the world for high-frequency communications work. However, there has always been a stormy controversy between Bell Labs and RCA about who really invented the antenna.

Edmund Laport (**Photo C**) wrote a chapter about "Long-Wire Antennas" in Henry Jasik's book *Antenna Engineering Handbook*. Laport invented the offset rhombic antenna and was a world-renowned radio engineer at the RCA. During the 1920s, he worked as a radio operator at KDKF in New York City, handling medical message traffic. However, by 1924 Laport became interested in engineering. He was one of the earliest proponents of microwave relaying because he was involved with the Radio Technical Planning Board organized by the FCC for postwar frequency allocations.

In 1943, under Laport's sponsorship at RCA, the company developed the two kilo-mega-hertz (KMz) CW-1 microwave system. The system was initially installed by the Western Union Company on its triangular New York City to Washington, D.C. to Pittsburgh circuit in 1947. It was the *first* real microwave radio relay system for commercial service. Terrestrial microwave has vastly expanded since then and now carries television signals, voice, and data. The terrestrial microwave relay system was subsequently challenged by satellite communications.

While Laport was busy with microwaves, Lowry was involved with the development of the Multiple Unit Steerable Antenna (MUSA) system used to communicate across the Atlantic by steering the antenna take off angle to the ionosphere. Lowry also helped develop equipment to explore atmospheric reflections via radar measurements. Lowry studied centimeter radiowave propagation over line-of-sight paths and microwave transmission over the Earth's horizon. His research led to more studies involving parabolic reflectors and the design and erection of such antenna systems.

Shortly after World War II, Lowry became involved with the Bell Telephone Satellite Communication Program. He was part of the team that dealt with control-center consoles at Holmdel, which was associated with the Telstar and Project Echo Satellites placed into Earth orbit (**Photos D**, **E**, and **F**). The Project Echo Satellite was used to bounce President Eisenhower's Christmas message to the world into space and then passively back to Earth.

Bruce and Beck had engaged in an interesting study concerning a North Atlantic radio circuit with the cooperation of the British General Post Office. Using a rhombic antenna at the Holmdel site, they were able to steer the main radiation lobe at various elevation angles and change the take-off angle to the ionosphere. The techniques were very simple and involved changing or distorting the diamond shape of the rhombic's physical dimensions, which directly altered the vertical firing angle.

Harold Friis also developed a method for determining the various elevation angles at which radiowaves arrived, again using the North Atlantic radio circuit. With the coopera-

tion of the British General Post Office, Friis used different methods to investigate radiowave propagation phenomena.

The first of Friis' methods used two different vertical antennas, one of which exhibited high-angle radiation pattern characteristics. Simultaneous monitoring of the antennas revealed that the vertical antenna with the strongest signal strength or input voltage to the calibrated receiver indicated whether the radiowave was arriving at a high or at a low angle. This led to the term "space diversity," which is widely used today. To obtain space diversity, two different types of antennas are used to receive the wave front at various incoming arrival angles. RCA's communications division used a communicator to switch different antennas to the receiver input to compensate for different wave arrival angle. Nowadays, an electronic comparison is made of signals from different antennas and the switch is made electronically.

It was a short step from this crude determination method for finding wave-front elevation angles to differential pulse output and phase methods. The pulse and phase methods were more sophisticated and provided a more refined measurement. The phase method was adopted later for steering the main radiation lobe in antenna arrays.

These propagation studies showed that wave fading resulted from phase interference. As a result, the wave front on one path would travel a longer distance to reach the receiving antenna while a shorter distance was traveled on the other path. The multiple paths resulted in signal fades due to this interaction. However, Friis and C.B. Feldman recognized that proper phasing techniques would eliminate this effect if multiple antennas were used in an array. Consequently, the signal quality would be much improved over long-haul radiotelephony circuits.

At the Holmdel site, an experimental antenna array was built consisting of one row of six rhombic antennas. The rhombic row was constructed in a straight line and directed at England. It stretched three quarters of a mile. Experimentation with the phasing lines between the rhombics, and adjustment for proper lead and lagging phase angles, provided a 7- to 8-decibel signal-to-noise improvement over a single rhombic antenna. By changing the phasing between antennas, the main receiving lobe could be steered through different elevation angles.

The culmination of all this work was the Multiple Unit Steering Array (MUSA) used for receiving. The MUSA was built at the WOO Manahawkin receiver site in New Jersey. This rhombic array had 16 rhombics stretched out over two miles on the salt marsh located near Barnegat Bay (**Photo G**). Each antenna was fed with a specially built brass coaxial transmission line. The transmission lines were placed under inert gas pressure to protect them from the corrosive marsh environment. Even though these lines were installed in the 1930s, many of the original lines were still in use during 1981.

The new receiving system was calculated to yield between a 12- to 13-decibel gain over a single rhombic antenna. It worked so well over the North Atlantic circuit that, during World War II, President Roosevelt had a regularly scheduled radiotelephony conversation with Prime Minister Winston Churchill using the Multiple Unit Steering Array at Manahawkin, New Jersey. The MUSA system helped advance single and double sideband transmission modes.

Robert W. Friis became directly involved with shortwave radio transmission voice facilities. After he joined Bell Labs in the 1930s, Friis helped develop high-power single and double sideband modulation modes later used at WOO, along with point-to-point and ship-to-shore radio telephony service.

Friis' 42-year career included research into and development of techniques for using ultra high frequencies (UHF) for radiotelephone. He contributed to the development of voice facilities between Provincetown and Green Harbor,

Photo H. One of the first microwave antennas and relay towers used across the nation. The antenna towers were capable of receiving and relaying messages over several hundred channels and television signals. (Courtesy of Bell Telephone Laboratories)
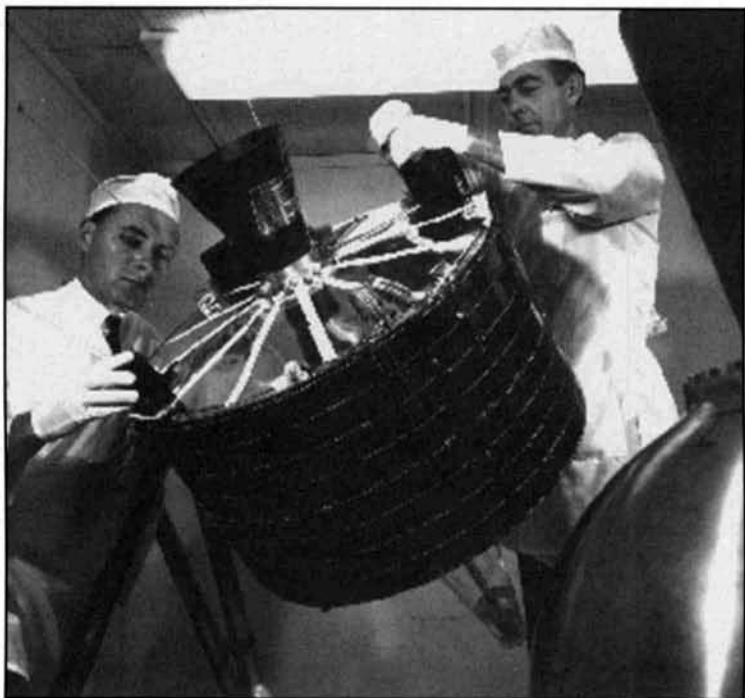
Photo I. Syncom satellite. (Courtesy of NASA)

Massachusetts. One of his other accomplishments was the development of high-frequency multi-channel radiotelephone circuits across the Chesapeake Bay. As World War II approached, Friis was a part of the fire-control radar development team. This team designed three different types of fire control radars for the Army and Navy. Friis further developed electronic-countermeasures equipment for the Office of Scientific Research and Development.

After the war, AT&T foresaw the need for communications growth and facilities using microwave radio. The company built a microwave relay system between Boston and New York (**Photo H**). Friis was one of the engineers who helped to design the microwave relay system. This successful system design was the basis for the construction of a nation-wide transcontinental microwave relay system. The microwave system would transmit telephone calls and network television signals from tower to tower across the U.S.

Friis was promoted to department head overseeing the Microwave Physical Design Department at Bell Labs. Designs from his department helped in the building of two new second-generation solid-state microwave systems operating on 4 and 6 gigahertz frequencies.

The Japanese

In Japan, the Meiji restoration took power from the Tokugawa clan in 1868 and forced the Land of the Rising Sun to become an industrial

and world power. After winning a foothold from the Russians on the East Asia mainland at Port Arthur, the Japanese were recognized by the world as a force in world politics. Since the Meiji restoration, Japan had modeled its educational and military system after the Europeans, in particular, the British.

Just north of Tokyo, along the coast, is Sendai, the city where Tohoku University is located. At the School of Electrical Engineering, Sintaro Uda had been working with parasitic elements using the discrete elements as wave reflectors and directors along with a driven dipole element. This particular technique displayed increased forward gain. Uda's colleague at Tohoku University, Hidetsugu Yagi, helped write a scientific paper on the phenomenon in 1926. Yagi was later recognized by the western world as the antenna's inventor. As a result, the antenna is frequently called a Yagi beam antenna today.

Although this invention had been around for a long time, nobody paid much attention to it. Americans George Brown and Wheeler Nagy explored the possibilities using parasitic elements in vertical arrays for the broadcast band and ultra high frequencies band (UHF). However, these studies never gained the necessary momentum for development until World War II.

Toward the end of the war, Imperial Japanese Navy (IJN) ships began deployment of the Yagi antenna in conjunction with radar. Soon this antenna was used exclusively as the radar antenna. The Americans had incorporated the Yagi antenna into American radar at the beginning of the war.

The Russians

Three Russians made important contributions to radio: Boris Grigor'evich Galerkin, Sergi Alexander Schelkunoff, and Aleksandr Aleksandrovich Pistol'kors. Despite the desperate situations created by Czarist rule and the Russian revolutions, these men rose to great heights.

Galerkin, who was arrested and sentenced to prison for participating in revolutionary activities, taught after his release in 1909. His field of expertise, as a Soviet engineer after the revolution, was in structural mechanics and elasticity. His theories and methods introduced a mathematical analysis into research pertaining to structures and machines. The numerical methods he devised efficiently and accurately approximated integration of equations used in elasticity theory.

Today Galerkin's methods are applied to antennas, making it possible to determine accurately the feed-point impedance using computers. Syracuse University Professor Roger F.

Harrington was responsible for introducing the Method of Moments, an outgrowth of Galerkin's methods as applied to antennas, to his classes. He also wrote a book about evaluating integrals employing matrix algebra methods. Improving upon Galerkin's methods, Professor Harrington revolutionized antenna design. Now, engineers can use their own table-top computer work stations to model antennas very rapidly using the work of Galerkin and Harrington.

Sergi Alekander Schelkunoff was instrumental in the hardware and theoretical microwave developments familiar to us today. After serving time in the Russian Army during World War I and the Russian Revolution, Schelkunoff journeyed to Japan and then to the U.S. Pacific Northwest.

Upon arriving in the U.S., Schelkunoff enrolled at Washington State College in 1922. He didn't speak English and had to work to support himself while attending college. However, Schelkunoff was an outstanding student and graduated in 1923 with a Bachelor's and a Master's degree in Mathematics. In 1928, he received his Doctorate in mathematics from Columbia University.

Later, while working for Bell Telephone Laboratories, Schelkunoff contributed to the fields of electromagnetic theory through theoretical studies of wave guides used in the transmission of microwaves. He was also involved in basic research on antennas for radio communications. During the 1930s, Schelkunoff solved the perplexing problem of cylinder antenna feed-point impedance. This pertained to thick dipoles that he visualized as behaving similarly to transmission lines. The equations he derived are still widely used today in the broadcast industry to solve problems concerning vertical antenna feed-point impedances.

Schelkunoff investigated wave-guide transmission line modes and discovered transmission line mode TE01. This discovery made it possible to use microwave radio for long distances. During World War II, he was a consultant on wave propagation at the U.S. Naval Station at San Diego, California. At the same time, he was also a member of the National Defense Research Committee. Schelkunoff wrote many technical papers and four books about antennas and theory. In addition, he received 15 patents pertaining to antennas, resonators, and wave-guides.

In 1927, as Schelkunoff completed his doctoral work, Alekander Pistol'kors graduated from the Moscow's Technical School. Pistol'kors began working at the Novgorad Radio Laboratories, where he formulated the solution to mutual impedance and antenna radiation resistance using the method of induced electromagnetic force. He outlined this method in a technical paper published by the *Institute of Radio Engineers Proceedings*.

During his career Pistol'kors developed many new antenna types and was noted for formulating a method of phase telegraphy. He also developed methods to design antennas from desired radiation pattern schemes. He was a corresponding member of the Academy of Sciences and a member of the Communist Party of the Soviet Union (CPSU). Pistol'kors was awarded the Order of Lenin, two other Russian state orders, and several medals.

Other important pioneers

Dr. Marion Gray joined Bell Labs in 1934 and pursued her interest in applied mathematics. She was interested in solving difficult problems related to the theory of wave-guides and horn antennas. At the same time, she worked on the mathematical theory of antennas. What's significant about Gray's work is that she solved the problems analytically because there were no computers to perform the tedious calculations.

The 1930s were a time of keen interest among engineers regarding antennas and their problems, and great leaps in knowledge were made. It was also a time when people such as Dr. George Brown did considerable research on the vertical antenna. Brown was responsible for various antenna analyses and wrote numerous papers describing ideas and theories. Brown described radiation patterns and current distribution using different shaped vertical antennas for use in the broadcast band. He also illustrated how to increase radiation resistance values using capacitive top loading and how to sectionalize a vertical antenna using inductive loading.

Photo J. Aft view of the *USNS Kingsport* AG-164. Notice the satellite disk radome behind the ship's house and stack.

Photo K. Forward view of the *USNS Kingsport* AG-164. Notice the two high-frequency discone cages on the forward decks. Directly behind is another satellite tracking antenna, composed of helical antenna array mounted on one post forward of the ship's house. Aft, on the ship's house, are more 35-foot whip antennas for high-frequency communications.

Brown's research wasn't confined to antennas and radiation theory. He explored radio-frequency heating, matching networks, and filters. He worked with electrical instrumentation and medical electronics and also studied the effects electric current had on penicillin dehydration and pasteurization.

Philip S. Carter worked at RCA's Rocky Point Laboratories on Long Island. A 1918 Stanford University graduate, Carter studied mechanical engineering. His contributions to radio were many. Carter wrote a paper on mutual impedance often used as a reference today. He is also the inventor of the folded dipole. One of Carter's awards, the "Citation for Overseas Work on Countermeasures V2 Rocket," was presented to him by the Secretary of War.

Microwaves and radar

World War II spurred advancement in the area of microwave frequencies. Radar systems were the motivating factor propelling radio engineering into the study, development, and use of centimeter wave lengths. Smaller and more advanced antenna systems were constructed for the new microwave frequencies. Antenna design progressed, providing good radiation pattern resolution. Most of this work was performed at the Massachusetts Institute of Technology and at Bell Telephone Laboratories. However, it was a young engineer named Phillip Smith at Bell who found a new way to do impedance calculations.

Phillip H. Smith came to Bell Labs directly after graduation from Tufts University in 1928.

During the Lab's research into overseas radio-telephone circuits, Smith studied antennas and transmission lines that carried radio energy to the antenna. He formulated a chart (the Smith Chart) using circles and coordinates to make transmission-line impedance calculations easier.

Smith and Morrison are responsible for the studies relating to shunt-feeding grounded vertical towers. However, Smith's expertise started with the development of commercial broadcast band antennas. This foundation prepared the way for his contributions to the design and development of certain types of radar antennas. For example, Smith was responsible for inventing the Cloverleaf antenna.

The next step in antenna development, once feed-point impedance could be determined, was matching the antennas to transmission lines. Matching impedance is required if efficient energy transfer is to occur between the transmitter and the antenna. Impedance matching can be considered stepping up or stepping down impedance to a desired value. In 1932, William Everitt wrote a fine treatise on the subject, which is found in his book *Communications Engineering*. While at the University of Ohio, he invented the pi-network known as a Collins filter. While antenna design went forward, so did transmission-line theory.

Matching networks had a disadvantage in that they were physically large at low frequencies and needed to be tuned to match a particular impedance. However, these same networks could be helpful in certain situations requiring phase advancement or phase delays, when they were applied to directional antenna arrays. Although these networks did serve their purpose and are still in use, the future would bring improvements.

In the *Proceedings of the Institute of Radio Engineers*, published in August 1959, Bell Labs scientist Ruthroff introduced his technical paper "Some Broadband Transformers." This paper presented a newer method of impedance transformation. Ruthroff used ferrite toroids and transmission line windings on the toroids to transform impedance over a wide bandwidth. Later, Dr. Jerry Sevick, W2FMI, developed applications pertaining to amateur radio antennas from Ruthoff's discovery. Ruthoff's methods have gained widespread acceptance and are in use worldwide.

Post WWII advancements

After World War II, more significant antenna advancements occurred as a result of technological developments in related areas. During the 1950s, the U.S. Air Force involved the University of Illinois in antenna research. The Air Force wanted to resolve the problem caused

by transmitting with many antenna systems on different frequencies. It wanted one antenna to do the whole job. The result was an antenna system with a broad frequency range. Naturally, it had a frequency response of 10:1. This antenna was sort of a modified Yagi-Uda antenna, but each individual dipole in the array was driven 180 degrees out of phase. The new antenna system was also scaled geometrically. This antenna is now known as the Log Periodic Array (LPA).

Despite its good points, the LPA didn't solve the problem of how one builds an efficient antenna that's physically small compared to its operational wavelength for very low frequencies. It's necessary to remember that the lower the frequency, the more difficult it is to build an antenna of the correct physical proportions to maintain resonance. Low-frequency antennas are usually fractional in physical dimensions compared to a corresponding wavelength.

In 1972, a remarkable new invention was patented in the U.S. by Joseph M. Boyer, W6UYH. He called his creation a tunable radio antenna. Boyer was able to see antenna top loading in a much different way. He visualized a transmission line loading a short vertical antenna. Then, by adjusting the transmission line length with a variable capacitor, the antenna could be tuned to any number of frequencies.

Articles written by Mr. Boyer appeared in *Electronics* magazine in February 1963 and in *73* magazine in September 1976. In his articles, Boyer presented and fully explained the antenna's operation. He also changed the publicly ascribed name from the Hula-Hoop to the antenna's real name "Directly Driven Ring Radiator Antenna."

Geostationary satellites

While advances were being made at high and very low frequencies, the same was true of microwave technology. During World War II, some radars were built by the Western Electric Company, the manufacturing arm of the Bell System. Western Electric built half the radar units used by the U.S. military. Their expertise in radar and microwave fields led to the development of the Nike Missile program, the Dew-Line, and the anti-ballistic missile systems. However, at the time, AT&T was in the best position to act upon a proposal made by Royal Air Force Officer Arthur C. Clarke.

In his article "Extra-Terrestrial Relays: Can Rocket Stations Give World Wide Radio Coverage?" published by *Wireless World* in October 1945, Clarke suggested the use of Earth satellites placed in geo-stationary Earth orbits as space-based repeaters or relay stations. These satellite repeaters would be equally spaced around the Earth's equator and would

use microwave frequencies for voice communications between continents and radio links between each satellite in orbit. Clarke was 20 years ahead of his time. The idea was sound; however, the rocket technology was not yet advanced enough to handle the payload weight. Moreover, there was another major problem with radio transmissions: weak signals in combination with noise generated in the atmosphere and within the receiver noise itself.

These problems were researched at Bell Labs, and the thermal noise problem was fairly well understood. The development of Microwave Amplification by Stimulated Emission of Radiation (MASER) and the horn antenna solved the problems associated with satellite communications. Scientists experimented with amplifiers and horn antennas in project Echo, which ushered in a new era in communications. The scientists used a ruby traveling-wave MASER with the horn antenna to receive weak signals reflected off the Echo-Two balloon satellite that orbited Earth in January 1964. The satellite had a 135-foot diameter and weighed 600 pounds. Its surface was constructed out of plastic and aluminum foil.

As a measure of how sensitive the system was, the noise temperature was 33 degrees Kelvin with the antenna pointed 10 degrees above the horizon. By comparison, the average microwave receiver was operated at a 1000-degree Kelvin noise temperature. The limiting factor was the thermal noise generated in the Earth's atmosphere and in the microwave receivers.

From the studies made at Bell Labs during the Earth satellite program that measured atmospheric noise, Arno Penzias and Robert W. Wilson detected a certain amount of background noise. They couldn't explain the noise, nor could they locate its source. After many measurements of other noise sources, a theory was advanced that researchers believed explained the observed phenomenon. The Bell Labs scientists thought it was due to the expanding cosmos. They also thought it could be the cosmos cooling down from its original high temperature state. They believed the noise must be thermal radiation still existing in the background. The discovery seemed another confirmation of the "Big Bang Theory" first proposed by physicists R.H. Dicke, P.J. Peebles, P.G. Roll, and D.T. Wilkinson, at Princeton University in New Jersey.

The National Aeronautical and Space Administration (NASA), in cooperation with the RCA, entered the space satellite game with the Synchronous Communications (Syncom) Satellite Program (**Photo I**). Syncom One was launched on February 14, 1963; however, the satellite's electronics failed on its way to geostationary orbit. Syncom Two was launched by

a Thor Delta 20 rocket on July 26, 1963, becoming the first operational geo-synchronous telecommunications satellite. Mr. Gunnar Pedersen inaugurated a telephone conversation from Geneva, Switzerland, to the United Nations in New York City using this satellite. On August 19, 1964 Syncom Three was successfully launched and reached geo-synchronous/geo-stationary orbit around the equator at a height of 35,900 kilometers.

To properly evaluate the Syncom series satellites around the world, NASA obtained the *Kingsport Victory* (Hull Number T-AK-239) from the U.S. Navy through the Military Sea Transportation Service. The ship sailed to Portland, Oregon, where she underwent conversion to a satellite communication ship by the Willamette Iron & Steel Company. On November 14, 1961, she was renamed the *USNS Kingsport* and reclassified as Hull Number AG 164.

The U.S. Army Satellite Communications Agency laid out the new design for the Kingsport, which underwent extensive alterations (**Photos J** and **K**). The ship was to provide support and conduct studies for the U.S. Defense satellite communications programs. A special high-frequency radio station was installed for ship-to-shore communications. The ship was also equipped with advanced tracking and telemetry equipment as well as anti-roll stabilization tanks. Along with these modifications, she received a 30-foot, gyro-stabilized, computer-oriented, triaxial, parabolic dish antenna. The antenna system was enclosed in a 53-foot, plastic, air-pressurized radome. The antenna could track high-altitude satellites at any vertical angle above the horizon with great precision.

The *Kingsport* left port on January 5, 1963, and sailed to Lagos, Nigeria, arriving January 18, 1963. From Lagos, the ship sailed to the Gulf of Guinea. On June 17 she returned to Lagos, remaining there until July 1, 1963, to conduct more tests and studies with Syncom Two. Before sailing to Lagos, the ship was painted white to remove any traces of her Navy gray hull in order to assuage fears in the civilian population that she was a warship.

In 1964, the *Kingsport* sailed to San Diego, California, and then on to Pearl Harbor, Oahu, arriving on March 25. Over the next 10 months, the ship spent time in the Western Pacific visiting Apra Harbor, Guam, and conducting communication experiments there. She then moved on to the Indian Ocean, evaluating Syncom Three. The ship also was able to support NASA's Gemini Program because of the vast wealth of knowledge gained from experiments made during the many months at sea.

The space age led to many spin-offs—not only in the area of communication satellites,

but in digital communications. Solid-state and miniature electronics were primarily responsible for the development of the information age. With compact devices, it was possible to convert analog data to digital data streams. This was the key. Digital transmission, also called telemetry, provided the means for the revolution in communications that continues today in the fields of fiber optics, metallic circuits (i.e., T1 Carrier), radio links, and software.

Communications and antenna systems are now entering a new era, far removed from their roots. Researchers have moved from studying centimeter wavelengths to exploration of millimeter wavelengths. It may be that the last great frontier will involve conquering the problems associated with microwave and satellite antenna design by stretching past the limits of current engineering techniques into the future.

Acknowledgments

BIBLIOGRAPHY
1. *Great Soviet Encyclopedia*, Moscow, USSR, 1974.
2. Mary Texana Loomis, *Radio Theory and Operating*, Loomis Publishing Company, Washington, D.C., 1927.
3. "What Ever Happened to ELF," *Spectrum*, Institute of Electrical, Electronic Engineers, June 1981, p. 22.
4. M. Harold Fredman and Jack Rosenblatt, *Direct and Alternating Current Machinery*, McGraw-Hill, 1963, p. 19.
5. B. Myril Reed, *Alternating Current Circuit Theory*, Harper and Brothers, 1948.
6. E.F.W. Alexanderson, "The Electrical Plant of Transoceanic Radio Telegraphy," *AIEE Journal*, Volume 42, pp. 693–703.
7. Mr. Wolfgang Schroder, personal correspondence and interview in Germany.
8. Mr. Charles Beulow, former head rigger and powerhouse engineer for the German Company, personal interviews.
9. "The Goldschmidt Radio Tower at Tuckerton, *Electrical Experimenter Magazine*, February 1914, p. 154.
10. Roy M. Nunn, *The Goldschmidt Wireless of Tuckerton, New Jersey*, Albright College, Reading, May 1967.
11. "The Goldschmidt Radio Tower at Tuckerton," *The Electrical Experimenter Magazine*, February 1914, p. 136.
12. Clinton B. DeSoto, *Two Hundred Meters and Down*, American Radio Relay League, West Hartford Connect, 1936, pp. 44–45.
13. "Lusitania Sunk on May 7, 1915, *The New York Times*, May 8, 1915.
14. Professor Dr. Walter Dieminger, Emeritus, Wissenschaftl. Mitglied Des Max-Planck-Instituts Für Aeronomie, DL6DS, personal interview about Professor Dr. J. Zenneck in Germany.
15. Jonathan Zenneck von Walter Dieminger, Deutsches Museum, Abhandlungen Und Berichte, 29 Jahrgang 1061 Heft 1.
16. Haraden Pratt, "Nikola Tesla," *Proceeding of the IRE*, 1956, pp. 1106–1108.
17. Doctor Cornelis J. DeGroot, "The High Power Station at Malabar, Institute for Radio Engineers Proceedings, Volume 12, pp. 693–722.
18. *Sparks Journal*, Society of Wireless Pioneers, Inc. Volume 3, No. 3, 1980.
19. H. Winfield Sector, "Currents of Ultra High Frequency and Potential," *Electrical Experimenter*, February 1914, pp. 151–152.
20. Naval Historical Center, Navy Yard, Washington, D.C. 20374.
21. Navy Department Library, Building 220, Room 220, Washington D.C.
22. WOO Wanamaker Marconi Station, personal interviews with AT&T employees at Ocean Gate, New Jersey.
23. William Peck Banning, *Commercial Broadcasting Pioneer, the WEAF Experiment 1922–1926*, Harvard University Press, Cambridge, Massachusetts, 1946.
24. A.C. Beck, biographical material supplied by Bell Telephone Laboratories, October 1965.
25. Lewis R. Lowry, biographical material supplied by Bell Telephone Laboratories, October 1962.
26. Robert W. Friis, biographical material supplied by Bell Telephone Laboratories.
27. Sergei Alexander Schelkunoff, biographical material supplied by Bell Telephone Laboratories, June 14, 1960, and August 22, 1960.
28. Marion C. Gray, biographical material supplied by Bell Telephone Laboratories, June 1953.
29. George H. Brown, *IRW Directory*, p. 18.
30. Philip S. Carter, *Institute of Radio Engineers Yearbook*, 1949, p. 46.

Peter J. Bertini, *K1ZJH*
Senior Technical Editor
20 Patsun Road
Somers, Connecticut 06071

QUARTERLY REVIEW

Summer reading

Crystal ladder filters have been a popular topic since the late 1970s. In the UK, authors Hardcastle and Hawker popularized them in the pages of *Radio Communications*. Later, Wes Hayward's articles in *QST* and other magazines showed how the passband of a ladder filter could be easily calculated using simple equipment.

Up until this time, most amateur radio homebrewers used the more costly mechanical or crystal-lattice type filters. We were blessed with a seemingly endless supply of surplus filters. Remember those MHz Electronics surplus ads in *Ham Radio* magazine hawking surplus Atlas CW, LSB, and USB filters for bargain prices during the '70s? The German KVG line of high-quality lattice filters was popular with those builders blessed with deeper pockets. While sources for surplus or new high-quality filters are slowly drying up or becoming prohibitively expensive, homebrewers now have the option of designing and building high-performance lattice filters using inexpensive microprocessor crystals.

A look at *Ladder Crystal Filters*

Ladder Crystal Filters, John Pivnichny, N2DCH, is a compact source of information on these filters. With his years of experience writing on the topic for various amateur radio publications, Pivnichny has much to offer the reader.

The book's first two chapters discuss crystal and capacitor characteristics and the measurement of crystal parameters using simple test jigs and measurement techniques. Chapters 3 and 4 delve into the mysteries of lower sideband and upper sideband lattice filters—applications ideally suited to the non-symmetrical nature of ladder filters. Next, Chapters 5 through 9 deal with the design of tunable, CW, and lattice filters. Finally, Chapter 9 discusses BASIC program listings for various filter

designs, including 6-pole Chebyshev USB and LSB filters, and 6-pole Butterworth USB and LSB filters. The reference section, with 40 listings, covers virtually every ladder filter article published in the amateur radio press.

Although little new ground is broken in the book, it handily provides a convenient single source for information that otherwise would be scattered over two decades and several different publications.

Ladder Crystal Filters by John Pivnichny is published by MFJ Publishing, headquartered in Starkville, Mississippi.

Communications Receivers— A testimony to vacuum-tube receivers

Communications Receivers-the Vacuum Tube Era: 1932–1981 is a testimony to the golden age of vacuum tube receivers by Raymond S. Moore. Several introductory pages are devoted to the history of communications receivers, and the remainder of the book lists the manufacturers and their wares in alphabetical order.

The book is copiously illustrated; however, I was somewhat disappointed in the overall quality of the equipment photographs. Many of the illustrations are poor reproductions of early advertisement photos, which were poor to begin with, and lack detail and definition. Many of the front panels on the equipment pictured appear washed out or, on the other extreme, entirely black. This is probably understandable given the magnitude of the undertaking; yet by comparison I was awed by the stunning photography in Schiffer Publishing's *Radios by Hallicrafters* book reviewed a few issues back.[1]

Years of manufacture, frequency coverage, and IF (and, if applicable, IF filter type) are provided for each receiver, along with the num-

ber and types of tubes used. Amateur radio and general coverage receivers are shown, as well as examples of military receivers. I was quite pleased to note that some relatively rare and obscure manufacturers are discussed. A brief thumbnail sketch outlines the important history for most manufacturers shown.

The guidelines limit the selection to superheterodyne vacuum tube receivers manufactured in the U.S. with a BFO and continuous tuning. While writing this I review, I was listening to my Zenith M660A general coverage receiver. It covers from 150 kHz through 30 MHz, has an ANL, BFO, standby switch, and even an RF gain control and bandspread logging scale. The M660A isn't rare and was made in Chicago

during the 1960s. Naturally, it was the first receiver I attempted to find in this book; unfortunately, I had no success. A small point, but it made a lasting first impression.

For boat anchor enthusiasts, *Communications Receivers*, published by RSM Communications, provides a good reference for identifying an unknown receiver or for basic technical information on a set. However, serious collectors will probably be disappointed given the lack of in-depth historical information and poor-quality illustrations. ∎

REFERENCE
1. Peter J. Bertini, K1ZJH, "Quarterly Review: *Radios by Hallicrafters*," *Communications Quarterly*, Summer, 1998, page 86.

CORRECTIONS

Notes on "The Double Rectangle" by Dan Handelsman, N2DT, Winter 1999

Dear Editor:

Here are some corrections to the double rectangle article.

On **Table 1**, the rightmost column, the column headings should read Rect on the top line and Impedance on the second line.

On **Figure 9**, the heading on the left y-axis should read Impedance: center and near. The heading on the right y-axis should read Impedance: far.

Under length 115' and spacing zero on **Table 7**, the text in the middle should read "simple rectangle."

In the substance of the article, I would like to call attention to a modeling error, which changes some of the findings—but only to a small degree.

1. **Modeling error.** The modeling error involved the defined direction of the center vertical. This was opposite to that of the end verticals and led to what appeared to be a phase inversion among the three. The three verticals are actually in phase.

2. **Gains.** The calculated gains are lower by anywhere from 0.05 to 0.12 dB. This is consistent across the board and none of the conclusions in the article need to be modified. The good news is that the changes in gain are not as drastic on either side of the optimum range of heights.

As an illustration and using a worst case example, a 115' ADR gives a gain over a range of spacings (and commensurate heights) from 0.8 to 8.0' of 4.33 to 4.43 dBi. The range of gain, from minimum to maximum, at a given antenna length is even smaller as the antennas get longer. This allows for considerable "lee-

way" (as defined in the original article) at all lengths and allows for construction based on practical dimensions rather than aiming for any narrow range. The gains at any height and spacing are due to two opposing trends; the greater gain as the height and spacing decrease (due to mutual coupling) being offset by greater resistive losses due to increased currents. The large plateaus of maximum gain in the longer antennas are due to an equilibrium being reached by these two opposing trends.

3. **Areas of transformation and anomalies.** There are no differences between the new data and the old as to the occurrence of these areas; the region of change from ER to ADR and the region of change from ADR to SDR. The "stepping" or "sawtooth" effect is due to a combination of the wrong model and artifacts due to NEC and MININEC, which occur at extremely close spacings of radiating elements.

4. **Dimensions.** The optimum heights for peak gain are consistently about two to three feet higher than in the published data. But, as I said above, there isn't much variation in gain for a given length; just pick a length, a practical height above 35 feet, and shoot for a target impedance. More height means higher feedpoint impedances and vice versa. Tuning is done simply by moving the center radiator.

5. **Currents.** It is now clear that the antenna functions as a series of broadside radiators spaced at various distances. The phase angles in the original article are wrong, but the current ratios hold.

Lastly, I wish to thank Steve Brandt, N7VS, for calling to my attention yet another DR variant, the H-Double Bay Antenna published by N4PC in *CQ*, September 1995. This is a horizontally polarized symmetrical DR for 30 meters, which is fed at an end. Thanks also to Darrel Emerson, AA7FV.

Dan Handelsman, N2DT
Chappaqua, New York

Robert R. Brown, NM7M
504 Channel View Drive
Anacortes, Washington 98221

POWER COUPLING
ON 160 METERS

On the role of polarization

For operating frequencies close to the gyro-frequencies of ionospheric electrons in the geomagnetic field (0.8 to 1.6 MHz), the conventional ideas about HF propagation must give way to more the general magneto-ionic theory[1] of the ionosphere. Thus, instead of just considering simple RF waves radiating first from antennas being refracted and exciting electron-atom collisions and then with signals captured by a receiving antenna, it is necessary to think about the polarizations involved. In particular, the transfer of RF energy to and from the ionosphere depends on antenna polarization as well as the various forms of elliptically polarized waves, from circular to linear, which propagate in the ionosphere. The question arises as to how RF from antennas couples into those polarizations so 160-meter signals can progress through the ionosphere to DX locations.

Geometry

At frequencies like 1.8 MHz, ionospheric electrons are under the influence of both nearby neutral and ionic constituents. This influence exhibits itself through collisions and ion chemistry, the geomagnetic field in which the constituents are embedded, and by the forces it exerts on them in their motions. Thus, wave refraction in the ionosphere is determined by its index of refraction, which varies according to the electron density as well as collision frequency, the wave direction relative to the magnetic field, and the field strength.

When explored in detail, it is found that RF waves in the magneto-ionic medium are generally elliptically polarized, with electric fields that may be resolved into two components at right angles and with phase differences of 90
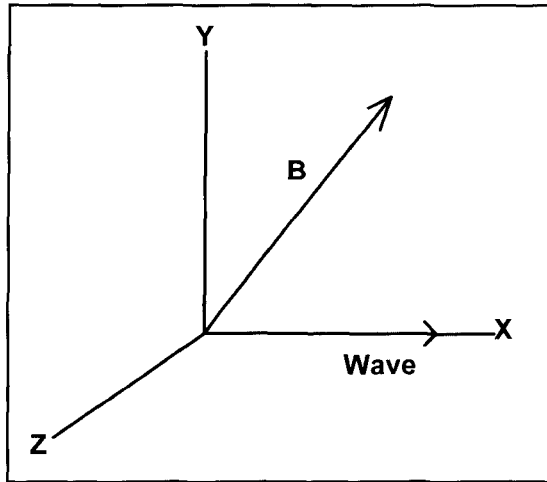
Figure 1. Right-handed coordinate system in the ionosphere with the x-axis in the path direction and the magnetic field B in the x-y plane.

degrees with each other. That sort of analysis is performed using a coordinate system in the ionosphere (**Figure 1**). There, the x-axis is in the direction of wave propagation. The direction of propagation and magnetic field B define a plane, and the y-axis lies in that plane and is perpendicular to the x-axis. A right-handed coordinate system in the ionosphere is completed with the addition of a z-axis perpendicular to the x- and y-axes. The electric field components of the wave are along the y- and z-axes.

While the electromagnetic analysis of signals is made in terms of the coordinate system at ionospheric heights, signals are launched, in both direction and elevation, from ground level and radio paths are reckoned using an earth-bound coordinate system. For propagation in the direction of the magnetic meridian, the relationship of the two coordinate systems is

Figure 2. Earth-frame and ionospheric frame of reference for a ray path above the horizon and in the magnetic meridian.

netic field. Thus, we have $X = (F_p/F)^2$, $Y = (F_h/F)$, and $Z = (v/2\pi F)$. Of interest to the present discussion are two other variables derived from the above: $Y_T = Y\sin(\theta)$ and $Y_L = Y\cos(\theta)$.

Wave propagation depends on those parameters, but special cases are important; for instance, when the geometry is such that the path is along the field direction or perpendicular to it. Those cases are termed longitudinal and transverse propagation. While waves in the magneto-ionic medium are generally elliptically polarized, the special cases represent limits where the wave polarization is either circular or linear. But more detail is found within those two cases; in particular, two types of elliptical polarization, the ordinary and extraordinary, which have different senses of rotation of the electric field. Thus, looking in the direction of the magnetic field, ordinary waves have a counterclockwise rotation of the electric field while the rotation of extraordinary waves is just the opposite—clockwise—and in the same sense as the rotation of electrons in the field.

The preceding remarks deal with the waves propagating in the magneto-ionic medium, and the index of refraction may be used to follow the waves along a path, say using the ray-tracing capability of the PropLab Pro program.[2] But it is important to consider not only ray paths but signal strengths, and to note that when the wave frequency is close to the electron gyro-frequency, as on the 160-meter band, the extraordinary wave is heavily absorbed by electron collisions.

Radio waves have their original polarization set by the transmitting antenna, but on entering the ionosphere they change and go over to the O- and X-waves that the ionosphere propagates. Just how the power in the incident wave divides between the O- and X-waves depends largely on the limit of the magneto-ionic polarizations at the bottom of the ionosphere—around 80 km altitude at night. The limiting polarizations, R, expressed as the ratio Ez/Ey of the electric fields that make up the elliptically polarized O- and X-waves, may be obtained from the general expression for wave polarization in magneto-ionic theory:

$$R = \frac{i}{2Y_L(1-X)}(Y_T^2 \pm \sqrt{(Y_T^4 + 4(1-X)^2 Y_L^2)})$$

(1)

by adding the limit $X = 0$; i.e., a negligible electron density at the bottom of the ionosphere. In that expression, the upper sign is for the O-wave and the lower sign for the X-wave.

For amateur radio operations on the 160-meter band, only the O-waves need be considered as the X-waves are heavily absorbed and

shown in **Figure 2**, where the Earth's curvature has been omitted. The x',y',z' coordinate system is fixed on the Earth, but, in general, the x,y,z system in the ionosphere changes orientation as signals launched from the ground vary in azimuth and elevation.

The electric field of a transmitted signal depends on the type of antenna used, its radiation pattern, and its surroundings. For this discussion, assume that either a vertical antenna is located at the origin of the x',y',z' coordinate system or that a horizontal antenna is oriented to have maximum radiation in the path direction. Thus, in **Figure 2**, the electric field from the ground-based antenna is perpendicular to the direction of the ray-path and in the y'-z' plane for a vertical antenna or along the x'-axis for a horizontal antenna. The components of the E-field vary in the fixed x',y',z' coordinate system for other azimuths and elevations, and the components in the x,y,z system vary, too, particularly because that system is not fixed in orientation and changes with signal azimuth and elevation.

Magneto-ionic theory

In magneto-ionic theory, the state of the ionosphere at a given point is described in terms of three parameters: X, Y, Z. These parameters involve the electron density, the magnetic field strength, the electron collisions with neutral constituents and are given in terms of frequencies (in MHz). The plasma frequency is F_p, the electron gyro-frequency is F_h, and the collision frequency is v; all are relative to the operating frequency. There is also the path-field angle θ—a geometrical feature related to the azimuth and the elevation of the path as well as the dip and declination of the geomag-

Figure 3. Ellipse showing the locus of the E-field for an O-wave advancing into the plane of the figure.

do not survive on long DX paths. So, turning to just the O-waves, the electric field rotates in the counterclockwise direction when viewed in the direction of magnetic field (**Figure 3**), and the limiting polarization for O-waves at the bottom of the ionosphere is given in terms of the parameters listed above:

$$R = \frac{i}{2Y_L} (Y_T^2 \pm \sqrt{(Y_T^4 + 4Y_L^2)}) \tag{2}$$

For signals coming up from below, the limiting polarization where the waves enter the ionosphere varies with azimuth and elevation, largely because changes in the path-field angle θ affect Y_T and Y_L. Such changes determine the extent to which power radiated by an antenna couples into O-waves in the ionosphere and is propagated toward DX locations at the different azimuths. The remaining fraction is coupled in as X-waves and lost by absorption.

Power coupling factor

Ionospheric propagation in the MF range (0.3 to 3.0 MHz) is important in radio broadcasting and has received considerable theoretical attention. One such study was conducted by staff members at the British Broadcasting Corporation.[3] Their work developed general results for the coupling of elliptically polarized waves going into and exiting from the lower ionosphere, and made applications at a radiation angle of 15 degreess. In addition, their work dealt with losses from changes in polarization when intermediate surface reflections occur over ground and sea water.

The study results may be applied to the question of power coupling into the ionosphere on the 160-meter band, but it is simply not feasible to extend the calculations to general DXing. This is because the number of hops can be quite

large, say up to 9 or 10 on very long paths, and the number of headings of interest to DXers are too numerous—requiring an inordinate amount of calculation time. Also, such calculations would be uncertain given the degree of variability of the ionosphere with solar cycles and the influence of magnetic activity on ionospheric properties within a given solar cycle. That being the case, the present calculations focus on the factors that are more stable, even in the face of solar and magnetic activity, yet affect the power coupling to the ionosphere and how the results differ for antennas having vertical and horizontal polarization.

In making the present calculations, there were two aspects to analyze due to the two coordinate systems associated with the problem. Thus, paths of interest to the radio amateur in the Earth-frame must be transferred to the ionospheric coordinate system before the methods of Phillips and Knight[3] are used to work out the power coupling. In that second frame of reference, two elliptically polarized waves are used (see **Figure 4**). The ratios of minor to major axes are M_a and M_b, respectively, for the incident wave and the limiting polarization at the bottom of the ionosphere.

For the problem of interest on 160 meters, the incident wave is taken to be linearly polarized (M_a=0) and at an angle Ψ_a with the y-axis of the ionosphere frame. Because X-waves are heavily absorbed, the mode of propagation in the ionosphere is via O-waves, with $\Psi_b = 0$ and M_b determined largely by the path-field angle θ. For that circumstance, the work of Phillips and Knight[3] provides the power ratio F of the induced wave to the incident wave as

$$F = \frac{(\cos (\Psi_a)^2 + M_b^2 \sin (\Psi_a)^2)}{(1 + M_b^2)} \tag{3}$$

Before the coupling factor F at a given site can be determined, it is necessary to relate the signal radiated by the antenna in the x',y',z' frame to the signal incident in ionospheric or x,y,z frame of reference. The heading and radiation angle are used to locate the entry points for signals into the ionosphere. Then a curvature correction is applied to the radiation angle for the distance between the antenna site and the point of entry. Next, the path-field angle is obtained using the geomagnetic declination and dip from the International Geomagnetic Reference Field.[4] With the magnetic field strength and the path-field angle θ, both Y_T and Y_L are determined and provide the minor-major axis ratio M_b for the limiting polarization.

Finally, using the methods of vector analysis, a linearly polarized E-field in the Earth-frame is transformed to the ionospheric frame, and the
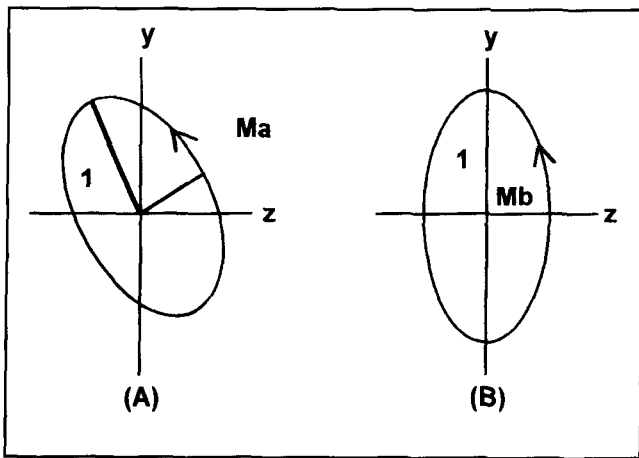
Figure 4. Polarization ellipses for the incident O-wave (A) and limiting polarization (B) at the bottom of the ionosphere, where M_a and M_b are the minor-major axis ratios, respectively.

power-coupling ratio F to O waves is obtained for the path. The present calculations, considering both vertical and horizontal polarization, were first performed for various points in the U.S., from the west coast to the midwest and then the east coast. Thirteen paths were used in the calculations, around the DX compass from Sweden (SM) to Siberia (UA0).

For 1.8-MHz signals at radiation angles around 15 degrees above the horizon, the power coupling ratios in the U.S. were relatively high—around 0.8 to 0.9—and only showed modest changes on the order of ±0.2 with azimuth variations for the case of vertical polarization. For the horizontal case, the power coupling ratios were low, the difference between unity and the value for the vertical case. Thus, horizontal power coupling ratios were on the order of 0.1 to 0.2 with changes with azimuth comparable to the case of vertical polarization. Note that for the case of horizontal polarization, it was assumed that the signals came from an antenna oriented broad-side to the azimuth of the path.

Results

The power coupling, F_v, to O-waves for signals from a vertical antenna is significantly greater than the factor, F_h, for a horizontal antenna at latitudes in the U.S. Note that the large vertical coupling factors in the U.S. result because the path-field angle is not far removed from 90 degrees; i.e., quasi-transverse propagation in the language of magneto-ionic theory. In that regard, the direction of the geomagnetic field lines is toward the north. Here, in the northwest, the International Geomagnetic Reference Field[4] indicates the Earth's field

points down into the ground at an angle of about 70 degrees from the horizontal, and the declination of the field is about 20 degrees east of north.

Looking at the orientation of the E-field from a vertical antenna relative to the magnetic field here in the northwest when the radiation angle is 15 degrees, one realizes that the two have the smallest angle (5 degrees) between them when the azimuth is to the north in the magnetic meridian (Figure 5). By the same token, the angle is greatest (35 degrees) when the azimuth is in the opposite direction and path-field angles for other azimuths lie between those two limits.

Note that when making calculations on the angle between the electric field and Earth's magnetic field, the actual situation is a bit different; the ray path meets the lower part of the ionosphere at a distance of about 275 km from the transmitting antenna. With the curvature of the Earth added to the calculation and the small changes in dip and declination from latitude and longitude changes with azimuth, the actual values differ from the simple situation described above. However the difference is not significant, so this discussion is adequate for purposes of illustration.

In this situation, vertically polarized signals heading in the various directions from here in the northwest have a large component of electric field along the direction of the local magnetic field. Thus, except for the large dip angle, those cases resemble the example cited in Davies.[1] Here we had a linear antenna in the N-S direction at the magnetic equator where E-W propagation produced only an O-wave because the RF E-field was parallel to the magnetic field. Consequently, for the Northwest, the coupling to O-waves should be the greatest for paths toward the north and least to the south. In that connection, calculations provide coupling factors of 0.88 toward Greenland (OX) or Siberia (UA0) while to the south, toward Easter Island (XR), the factor is not much different at 0.85.

As noted in Davies,[1] the orientation for the RF field relative to the geomagnetic field is the important thing when it comes to coupling RF into the ionosphere, not the orientation with respect to the Earth's surface. As a result, vertically polarized signals at latitudes where the dip angle is large produce similar degrees of signal coupling to the ionosphere just as for the case of horizontally polarized signals at latitudes where the dip angle is small.

The above results are for one corner of the U.S., and, while it is of interest to consider how the coupling factor changes with latitude and longitude within the U.S., large changes are not expected. The dip angles at other latitudes in
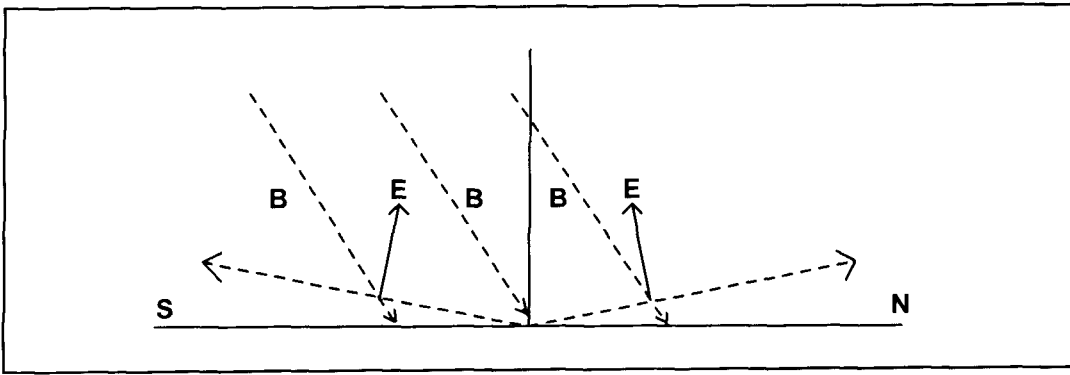
Figure 5. E-field orientations relative to the local magnetic field for paths in the magnetic meridian, heading to the north and the south.

the U.S. are comparable, say 66.5 degrees for Wichita, Kansas; 68.2 degrees Philadelphia, Pennsylvania; or 62.2 degrees for Dallas, Texas. Thus, the range of coupling factors for those locations were 0.87 to 0.82, 0.89 to 0.85, and 0.78 to 0.73, respectively.

In contrast to circumstances at mid-latitudes in the U.S., one can look at the power-coupling question at lower latitudes, using regions similar to those where DXpeditions were conducted recently: S21XX in Bangladesh (23.8N, 90.5E),

XZ1N in Burma (17.8N, 96,4E), VK9XY on Christmas Island (10.5S, 105.6E), and VK9CR on Cocos-Keeling Islands (12.1S, 96.9E). At those sites, the geomagnetic latitude is quite low and the Earth's field is close to being horizontal; i.e., parallel to the Earth's surface.

On that basis, one would expect strong coupling to O-waves if the RF electric field were parallel to the geomagnetic field. This brings up the difference between antennas with horizontal and vertical polarization in a most strik-
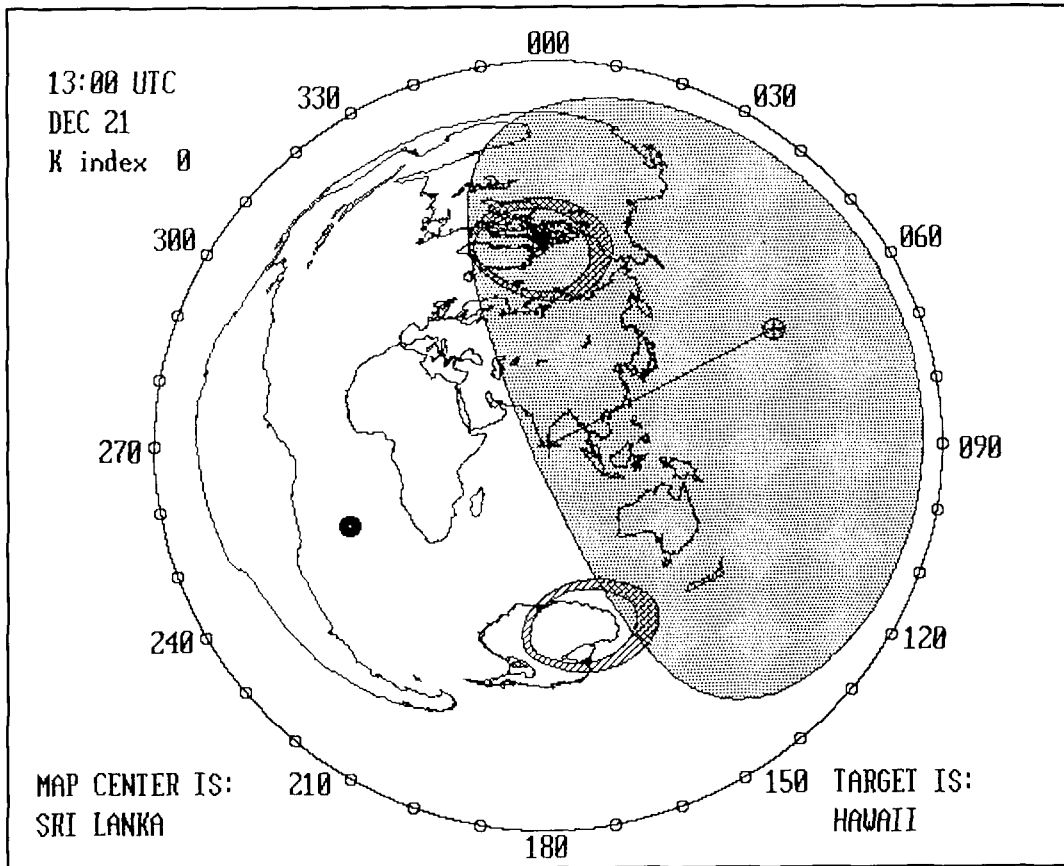
Figure 6. Azimuthal equidistant map centered on Sri Lanka.
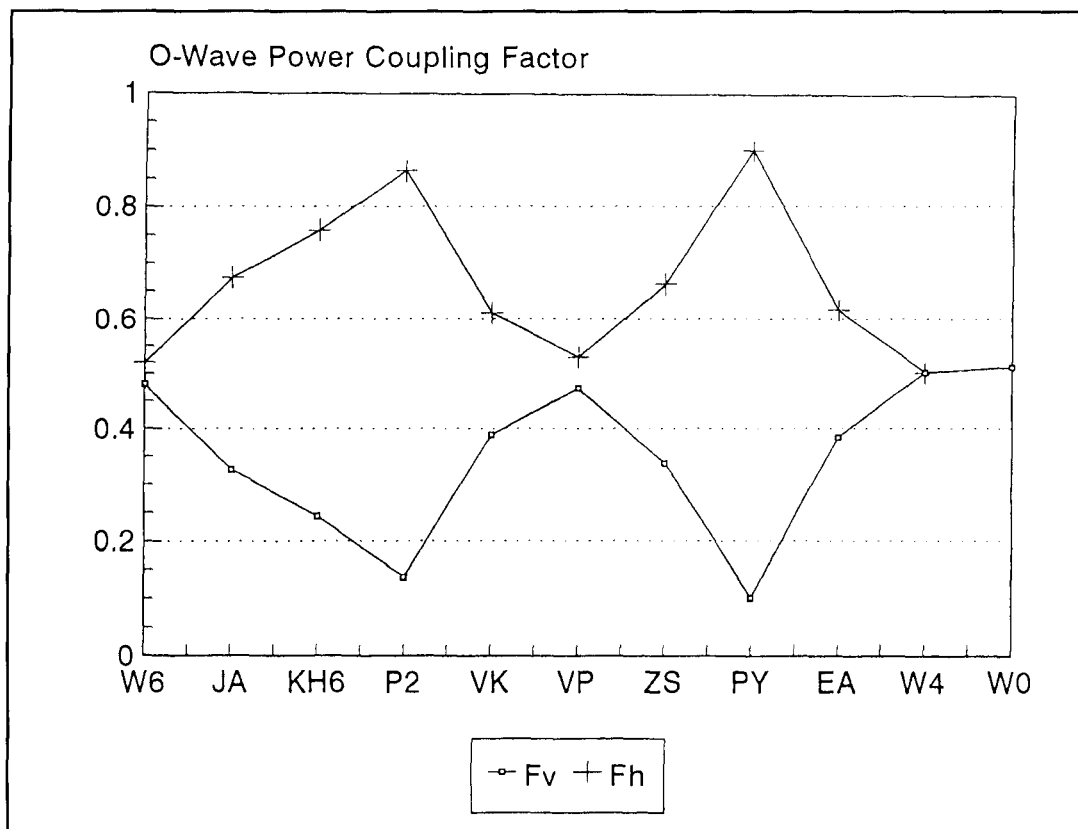
Figure 7. O-wave power coupling factors for DX paths from Sri Lanka and signals with vertical and horizontal polarizations.

ing fashion—as those azimuths where horizontal polarization has a strong coupling to the ionosphere are also where vertical polarization would have the weakest coupling. Considering that vertical antennas are preferred by many on 160 meters, this points to some potential problems for those who organize DXpeditions to low latitudes.

Consider 160-meter operations from Sri Lanka (4S7) at 7.4N, 80.4E, where the magnetic dip angle is only -2.4 degrees from the horizon and the declination is only 2.7 degrees west of north. There, the Earth's magnetic field is almost parallel to its surface and pointing northward from Sri Lanka. With that, one can look at paths from Sri Lanka and easily identify the directions to the East and West where the power coupling would be greatest for horizontal polarization or, on the other hand, the lowest for vertical polarization.

Using the azimuthal equidistant map centered on Sri Lanka, in **Figure 6**, one can expect strong coupling for signals with horizontal polarization going to the East from Sri Lanka, say toward Hawaii (KH6) or New Guinea (P2), or to the West, say toward Brazil (PY), and weaker coupling for signals going north or south, toward North America (W0 and W6) and the Antarctic Islands (VP8). This is born out by

calculations shown in **Figure 7**, where the power coupling for horizontal polarization reaches 0.8 to 0.9 toward New Guinea (P2) and Brazil (PY). In those directions, power coupling from vertical antennas is at a minimum.

Discussion

Power coupling into the ionosphere is independent of solar and magnetic activity. It depends only on radiation going through the neutral atmosphere toward the lower ionosphere, where the electron density is extremely low, and the static components of the geomagnetic field. On moving into the ionosphere, however, the elliptically polarized waves that then develop may be affected by solar and magnetic activity as they advance through the varying electron density and magnetic field in the various ionospheric regions.

In addition, the state of wave polarization changes because the reflections at the Earth's surface as the limiting polarization, when the radiation exits the ionosphere, will not be the same type of polarization on re-entry after reflection. This is due to amplitude and phase changes on reflection at the Earth's surface. In view of the obvious complexity, it is not practi-

cal to work out the polarization changes on any but the simplest propagation paths, as was done by Phillips and Knight.[3]

However, the power coupling into the ionosphere resulting from polarization effects can be an important factor in amateur radio communication. Thus, the fundamental magneto-ionic result cited in Davies[1]—efficient excitation of O-waves when the RF electric field is parallel to the magnetic field—means that the motions of ionospheric electrons excited by the wave E-field are parallel to the magnetic field. For that case, there is no magnetic force on the electrons and the situation is quite "ordinary" in that it is the same as that which would prevail in the absence of the geomagnetic field.

In any event, vertical antennas will be more effective on 160 meters at latitudes where the geomagnetic field has a strong vertical component than antennas with horizontal polarization. But that result assumes the horizontal electric field is perpendicular to the direction of propagation—in effect that the antenna has been positioned or steered for the most favorable polarization. Most horizontal antennas used by amateur operators do not meet that criterion and thus their radiation fields have both horizontal and vertical components, making it very difficult for any comparison with the results cited here.

However, for contesters or those making DXpeditions to sites at low magnetic latitudes like the one suggested for Sri Lanka, it is clear that horizontal antennas would be appropriate to complement vertical antennas. Both vertical and horizontal antennas are equally effective for paths to the north or south, where the limiting polarization is circular, but horizontal antennas are more effective for paths perpendicular to the declination of the local magnetic field where the limiting polarization is linear, as shown in **Figure 7**. ■

REFERENCES
1. K. Davies, *Ionospheric Radio*, Peter Peregrinus Ltd., London, U.K., 1989.
2. Cary Oler, "PropLab Pro: A High-Frequency Radio Propagation Laboratory," *Solar Terrestrial Dispatch*, 1994.
3. G.J. Phillips and P. Knight, "Effects of polarization on a medium-frequency sky-wave service, including the case of multihop paths," *Proceedings of the IEE*, Vol. 112, No. 1, January 1965.
4. IGRF 1995, *International Geomagnetic Reference Field, Utility Programs for Geomagnetic Fields*, National Geophysical Data Center, Boulder, Colorado, 1996.

PRODUCT INFORMATION

JTM Clamp Meters

A line of digital multimeters, including three clamp meters, is available from Jensen Tools. The JTM-30 is a miniature AC/DC current clamp meter that accepts up to 22-mm conductors. The JTM-33 is a true RMS AC-only meter. It has display hold and is four-foot-drop tested. The meter accepts up to 40-mm conductors. The JTM-39 is Jensen's top-of-the-line AC/DC true RMS meter with frequency measurement via the clamp. It can accept 51-mm conductors or 24 x 60-mm bus bars, and has also passed the four-foot-drop test. Each meter comes with test leads, a manual, and batteries. Jensen offers a three-year warranty on all JTM meters.
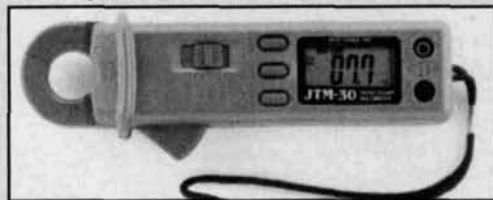
For a copy of the latest edition of the Jensen Tools Catalog, write to Jensen Tools at 7815 S. 46th Street, Phoenix, Arizona 85044 or call 1-800-426-1194.

Tektronix Oscilloscope Now Available with Four Channels

Tektronix, Inc., has introduced the TDS224 four-channel, real-time, 100-MHz digital storage oscilloscope (DSO) ass the newest member of the Tektronix TDS200 Series.

These 60- and 100-MHz units are designed for digital design engineers and technicians working with 8-bit micro-controllers. The TDS200 Series is also used in the manufacturing and service sectors.

The TDS200 Series oscilloscopes use Tektronix' oversampling technology. All three models—the TES210, TDS220, and the TDS224—sample at 1 GS/s, providing full bandwidth single-shot acquisition simultaneously on all channels.

For more information about the TDS224 and other Tektronix products, visit the company's home page on the World Wide Web at <http://www.tektronix.com> or call 1-800-426-2200.

A Practical Amateur Application of the Smith Chart

You just put up a new vertical antenna, but when you connect the antenna analyzer, there's no 50-ohm match. Now what?

Steve Sparks, N5SV

When I turned my 60-foot tower into a 160-meter folded monopole,[1,2] the 50-ohm match was terrible (see **Photo A**). I had the option of using a transmatch at the transmitter end of the coax, but that would leave a badly mismatched termination impedance at the far end of the coax. I thought, "Why not put the matching network at the antenna end of the coax and use the transmatch to move up and down the band?" Well, the L network recommended in **Reference 1** didn't work either.

Now what?

The solution was to begin at the antenna. First, I had to find a way to measure the complex impedance of the antenna over the frequencies of interest.[3] I wanted to use my folded monopole on the 160-meter band, 1.8 to 2.0 MHz. **Reference 3** gave several ways to measure these impedances. I used an impedance bridge, a signal generator, and a spectrum analyzer as the detector. One could also use a receiver as a detector with the bridge. **Table 1** gives the results of these measurements. Once I knew the antenna input impedance, I had to calculate a matching network to transform these impedances to 50 ohms.

The Smith Chart

The Smith Chart (**Figure 1**) provides an easy solution to the question of using a two-component L network to transform a complex impedance to a 50-ohm resistive load—a great match for the 50-ohm coax running from the transmitter to the antenna. **Figure 2** shows eight examples of a two-element L network. After you normalize your Smith Chart, pick the best solution network based on **Figure 2**. In my case (**Figure 1**), all normalized points made up a circle in the upper right quadrant corresponding to (a) in **Figure 2**. Z_O was the transmission-line end of the L network and Z_L the antenna end of the L network.

Let's look at a counter example. Consider (d) in **Figure 2**. Use a series capacitor to move to Z_1, which is the intersection of circle Z_L and semi-circle Z_1. Next, use a shunt inductor to move from Z_1 to Z_0. Now all we need to do is generate the values for X_C and X_L off of the Smith Chart (**Figure 1**). Pick a solution where the line length (i.e., movement around the chart) has some distance to it. If the line length is very short, component values will be very small and may be hard to construct.

I'll use my problem in **Figure 1** as an example for determining the values needed for shunt capacitance and series inductance as shown in **Figure 2A**.

Actual matching values generated

Starting at the antenna end, Z_L, draw a circle of constant impedance using a compass whose
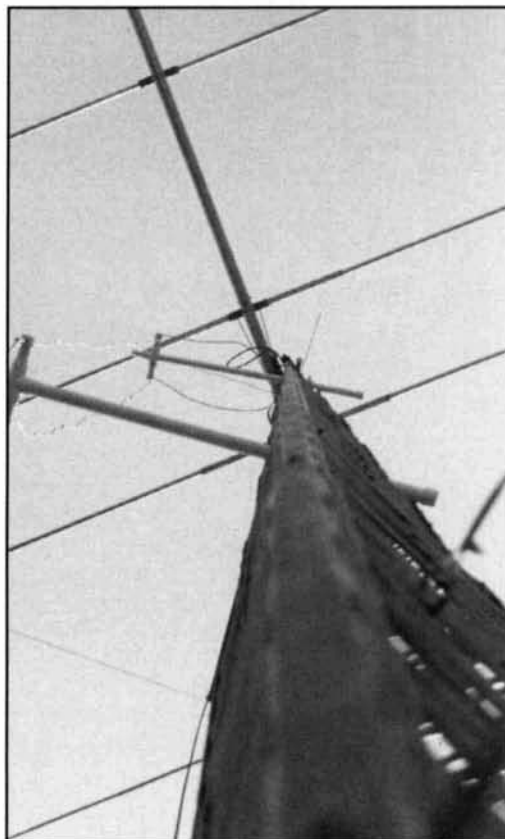
Photo A. Looking up the 60-foot tower with the three-wire "cage" held out from the tower two feet. The wires are tied together at the top (connected to tower) and bottom (connected to matching network). A Mosely 67-C-3 sits atop the tower.

Figure 1. Smith Chart.

Figure 2. Network impedance transformations.

Case 1	Case 2

Case 1

@ 1:1 SWR

Power = 1500 watts

Eq. 7. $\epsilon = \sqrt{P \times R}$

$\quad = \sqrt{1500 \times 50}$

$\quad = 274 \, V_{RMS}$

$\quad = 387 \, V_{RMS}$

Case 2

@ 1:1 SWR

$Z_{ANT} = 90 + j110$

$Z_{ANT}L \, 142 \, \Omega \quad \epsilon = I \times Z_{ANT}$

polar $\quad\quad I = \sqrt{P/_{R_R}}$

$\quad\quad\quad\quad \epsilon = \sqrt{1500/_{90}} \times 142$

$\quad\quad\quad\quad \epsilon = 4.08 \times 142$

$\quad\quad\quad\quad \epsilon = 580 \, RMS$

$\quad\quad\quad\quad\quad = 820 \, V \, peak$

Note: I is a function of the Real $\underline{R}_{ANT}$ not ZL

Case 3

@ 1:1 SWR

P = 1500 watts

$Z_{ANT} = 120 + j190$

C = 200 pF @ 3.65 MHz = 218 Ω

$\epsilon = I Z$

$I = \sqrt{P/_Z} = \sqrt{1500/_{225}} = 2.58 \, A$

$\epsilon = 2.58 \times 218$

$\quad = 562 \, V_{RMS}$

$\quad = 795 \, Vpeak$

Case 4

@ 1:1 SWR

Zo = 50 Ω

$I = \sqrt{P/_R}$

$\quad = \sqrt{\dfrac{1500}{50}} = 5.48 \, A$

C = 200 pF @ 3.65 MHz = 218 Ω

E = I Z

$\quad = 5.48 \times 218$

$\quad = 1194 \, V_{RMS}$

$\quad = 1689 \, Vpeak$

Figure 3. Techniques for calculating the L network capacitor voltage requirements for 1500 watts of power.

center is "0" on the far left of the horizontal resistance axis. The length of this drawn line is from the worst case (highest 1.9 MHz) to where it intersects the 1.0 constant resistance circle whose 1.0 value represents 50 ohms normalized on the Smith Chart.

X_C value/capacitance value

To calculate an X_C value, move the two end points to the left and opposite side of the Smith Chart onto a constant resistance circle where you can read the deltas of X_C normalized. You

can do this with a ruler. All points are equidistant from the Smith Chart's center. Looking at **Figure 1**, you can see that the 1.9-MHz point translates to a value of 0.23 as read from the susceptance circle(s). Also, note that the intercept point of the 1.0 constant resistance circle translates to a value of 0.26 in the upper portion of the Smith Chart. This left side of the Smith Chart represents susceptance or $1/Z_0$. To calculate the capacitance value, do the following:

1. Add the two deltas together (0.23 + 0.26 = 0.49) total line length.

Table 1. The complex impedance.

Frequency	A R±jX	B R±jx normalized
1.8	105+j333	2.1+j6.6
1.825	175+j318	3.5+j6.3
1.850	218+j216	4.3+j4.3
1.875	116+j150	2.3+j3.0
1.900	57+j199	1.1+j4.0
1.950	60+j279	1.2+j5.4
2.000	100+j310	2.0+j6.2

Table 1. **A**: The complex impedance as measured at the bottom of the wires running up the tower. **B**: The same impedance as **A**, but normalized to 50 ohms for plotting on the Smith Chart of Figure 1.

2. Divide 0.49 by 50 to get un-normalized susceptance (0.49/50 = 0.0098).

3. Divide 0.0098 into 1 to move from susceptance to reactance (1/0.0098) = 102 ohms. This is the X_C value.

4. Convert 102 ohms of X_C to capacitance at 1.9 MHz in picofarads:

$$CpF = 1/(2pi*F_{MHz}*X_C) =$$
$$1/(6.28*1.9*102) = 820 \text{ pF}$$

This is the value of the shunt capacitor per **Figure 2A**.

X_L value/inductor value

Starting at the antenna, we transformed the antenna impedance to the 1.0 (50 ohm) constant resistance circle. Now we must move the impedance to the 1.0- (50-ohm) center point, or the pure resistance point on the Smith Chart. Do this by adding just enough inductance to intercept that point. *How much inductance?* Well, if we read the Smith Chart, the center line is zero and our transformed worst case point intercept with the 1.0 circle reads approximately 5.3. This is normalized inductive reactance. To calculate the inductance required, do the following:

1. Multiply 5.3 by 50 to un-normalize. This equals 265 ohms reactance.

2. Convert 265 ohms X_L to inductance at 1.95 MHz.

$$L\mu H = X_L/(2*\Pi*F_{MHz}) =$$
$$265/(2*3.14*1.95) = 21.6 \text{ } \mu H$$

In my case, this inductor—about 2.5 inches in diameter and about 8 inches long—was a variable purchased at a swap meet.

Voltage ratings for L matching networks

Working with the inductor is easy. The voltage between each turn is quite small relative to the end-to-end voltage. Also, the ends are usually far apart and can withstand considerable voltage. The capacitor can present a problem. All the voltage is across the gaps between the plates; therefore, all of these gaps must withstand the total voltage.

Figure 3 provides four examples that cover all the cases of **Figure 2**. For the antenna described in this article, we would use Case 2. Looking at the values of **Table 1**, columns A and B, we see the antenna real impedance comes close to 50 ohms at about 1.890 MHz. Using this worst case value for the RF current, we calculate it to be 5.48 amps. This comes from the formula of Case 2 (**Figure 3**) where Imax = sqrt (1500/50). The complex antenna impedance at 1.890 MHz from **Table 1**, column A is about 50+j200. The square root of the sum of the squares yields 200 ohms in polar form. The voltage across the capacitor is: E = I x Z_{polar} or 5.48 x 200 = 1095. This is 1095 volts RMS and multiplying by the square root of 2 yields 1550 volts peak. We'll need a capacitor rated at about 2500 volts or better for our shunt variable capacitor.

I hope this article helps others calculate their way through matching networks. By the way, these same principles are at work in your trans-match/antenna-tuner, only at the other end of the coax. The antenna shown in **Photo A** was used in the 1998 CQ160-meter SSB Contest where I worked 49 states and 12 countries. This included OT8T, the famous Belgian contest station. I welcome questions and comments on this article. My e-mail address is <ssparks@wyle.com>. ∎

REFERENCES
1. John Devoldere, ON4UN, *Low Band Dxing*, American Radio Relay League, Newington, Connecticut.
2. Jeff Briggs, K1ZM, *Dxing on the EDGE*, American Radio Relay League, Newington, Connecticut.
3. Wilfred N. Caron, "The Hybrid Junction Admittance Bridge," *The Hybrid Junction Admittance Bridge, Volume 3*, American Radio Relay League, Newington, Connecticut.

BIBLIOGRAPHY
1. Application Note #154, Hewlett Packard Corporation.
2. Philip H. Smith, *Electronic Applications of the Smith Chart*, Noble Publishing.
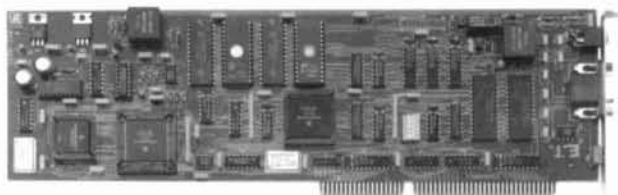
CLOVER-2000

High Performance HF Radio Protocol

Fast, Reliable, Economical Communications

PCI-4000/2K Internal Modem

DSP-4100/2K External Modem

CLOVER-2000 is an advanced modem waveform and protocol that is specifically designed for high rate data transmission via High Frequency (HF) radio. Available for use in HAL DSP-4100/2K or PCI-4000/2K modems, CLOVER-2000 automatically adapts to changing HF propogation conditions.

High Throughput & Adaptive ARQ:
CLOVER-2000 sends data over standard HF SSB radio channels at 3000 bps. Including error correction and ARQ overhead, CLOVER-2000 will deliver up to 2000 error-corrected bits/second over an HF radio link. The -50dB occupied bandwidth of the transmitted signal is only 2000 Hz. CLOVER-2000 may be used with *any* good-quality HF SSB transmitter and reciever.

The CLOVER demodulator measures Signal-to-Noise ratio (SNR), Phase Dispersion (PHS), and Error Corrector activity (ECC) of each data block received. This information is used to determine which of the five modulation formats should be used during the next 5.5 second transmission. In contrast, other adaptive systems use only 1 or 2 modulation formats and selection is based only upon errors as they are received and not on real time measurements of actual channel parameters.

Error Correction Coding:
CLOVER uses Reed-Solomon error correction coding to combat burst-errors that occur on typical HF transmissions. Other ARQ modes may not include in-block error correction or use formats that require long interleave times to combat burst errors.

Bi-directional ARQ:
The CLOVER ARQ protocol allows real-time adaptive transmission of data *in both directions* on the ARQ link *without the need for special "over"* commands. CLOVER is also data transparent; the modem will send any 8-bit stream provided without modification and without requiring special command sequences. CLOVER-2000 is a low-cost and high throughput solution to your HF data communications problems. It is the "waveform of choice" for thousands of users throughout the world.
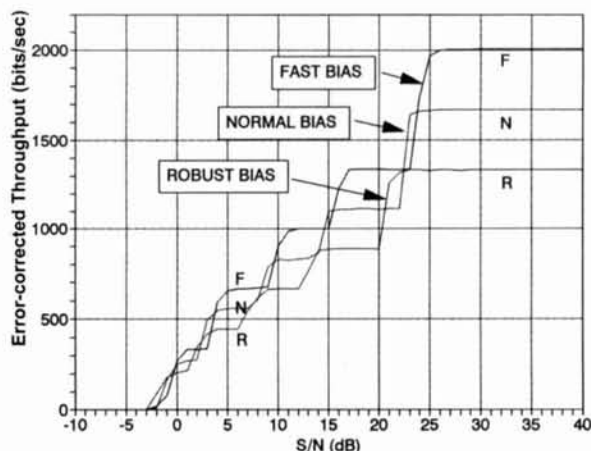
CLOVER-2000 Data Throughput
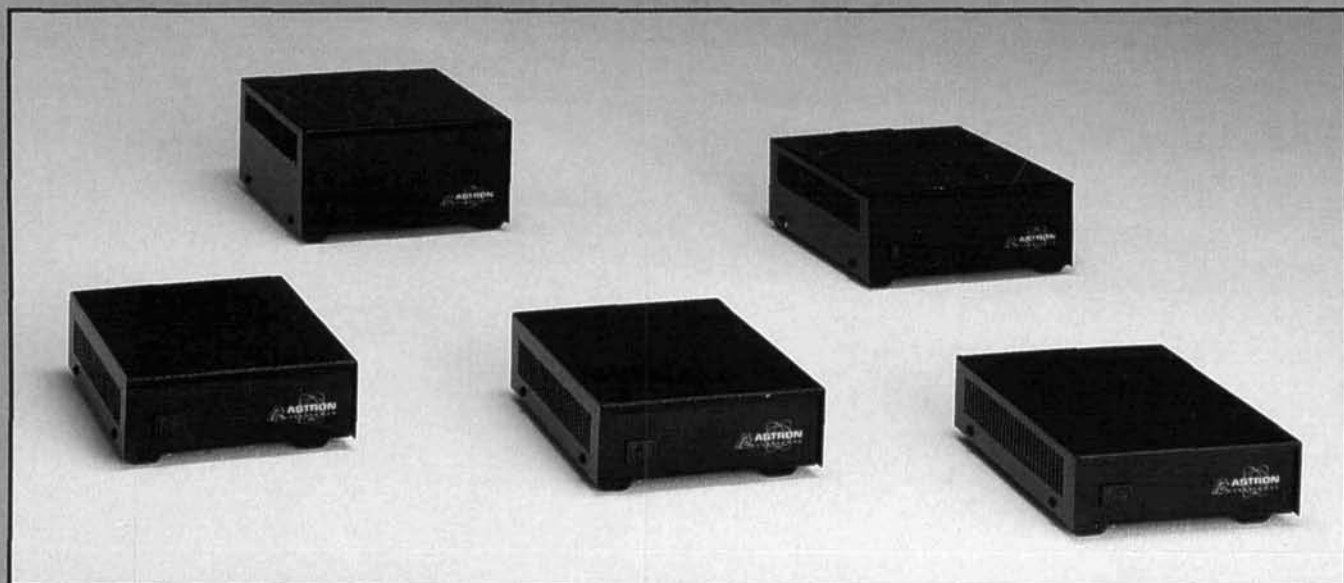
HAL COMMUNICATIONS CORP.
1201 W. Kenyon Road, P.O. Box 365
Urbana, Illinois 61801-0365
Phone: (217) 367-7373 FAX (217) 367-1701
www.halcomm.com