K1MC describes the "Ultimate Sidetone," a single-tone, keying-shaped sidetone circuit. Add it to a vintage receiver to monitor your CW sending.

TOKYO HY-POWER

HL-1.5KFX
HF/50MHz Linear Power Amplifier

NEW!

Auto Band Set

Specifications

Frequency:
1.8 ~ 28MHz all amateur bands including WARC bands and 50MHz

Mode:
SSB, CW, RTTY

RF Drive:
85W typ. (100W max.)

Output Power:
HF 1kW PEP max., 930W CW (typ.)
50MHz 650W PEP max.

Matching Transceivers for Auto Band Decoder:
Most modern ICOM, Yaesu, Kenwood

Drain Voltage:
53V (when no RF drive)

Drain Current:
40A max.

Input Impedance:
50 OHM (unbalanced)

Output Impedance:
50 OHM (unbalanced)

Final Transistor:
SD2933 x 4 (MOS FET by ST micro)

Circuit:
Class AB parallel push-pull

Cooling Method:
Forced Air Cooling

MPU:
PIC 18F452 x 2

Multi-Meter:
Output Power – Pf 1Kw
Drain Voltage – Vd 60V
Drain Current – Id 50A

Input/Output Connectors:
UHF SO-239

AC Power:
AC 230V (200/220/240V) – 10A max. (default)
AC 115V (100/110/124V) – 20A max.

AC Consumption:
1.9kVA max. when TX

Dimension:
10.7 x 5.6 x 14.3 inches (WxHxD)/272 x 142 x 363 mm

Weight:
Approx. 20kgs. or 45.5lbs.

Accessories Included:
AC Power Cord
Band Decoder Cables included for Kenwood, ICOM and Yaesu
Spare Fuses and Plugs
User Manual

Optional Items:
Auto Antenna Tuner (HC-1.5KAT)
External Cooling Fan (HXT-1.5KF for high duty cycle RTTY)

This compact and lightweight 1kW desktop HF/50MHz linear power amplifier has a maximum input power of 1.75kW. Our solid-state broadband power amp technology makes it **the smallest and lightest self-contained amplifier in the industry**.

Typical output power is 1kW PEP/SSB on HF and 650W on 6m band with the drive power of 85-90W. Bands set automatically with the **built-in band decoder**. You can forget about the band setting when the amplifier is connected to your modern radio through **supplied band data cables for ICOM CI-V, DC voltage (ICOM, Yaesu), and RS-232C (Kenwood)**. Manual band setting selectable as well.

All these data cables are included with the amplifier.

Features

- Lightest and most compact 1kW HF amplifier in the industry.
- The amplifier's decoder changes bands automatically with most ICOM, Kenwood, Yaesu.
- The amp utilizes an advanced 16 bit MPU (microprocessor) to run the various high speed protection circuits such as overdrive, high antenna SWR, DC overvoltage, band miss-set etc.
- Built in power supply.
- AC 230V (200/220/240V) default and AC 115V, (100/110/120V) (selectable).
- Equipped with a control cable connection socket, for the HC-1.5KAT, auto antenna tuner by Tokyo Hy-Power Labs.
- Two antenna ports selectable from front panel.
- Great for desktop or DXpedition!

Watch for our NEW PRODUCTS to be announced soon!

TOKYO HY-POWER

QEX

In order to ensure prompt delivery, we ask that you periodically check the address information on your mailing label. If you find any inaccuracies, please contact the Circulation Department immediately. Thank you for your assistance.

About the Cover

Mal Crawford, K1MC, describes his "Ultimate Sidetone," a single-tone, keying-shaped sidetone circuit. Mal added this circuit to Heath SB-301 and SB-303 receivers to monitor his CW sending. You can modify the circuit values to suit other receivers.

Features

Columns

Jan/Feb 2007 *QEX* Advertising Index

Empirical Outlook

Doug Smith, KF6DX
kf6dx@arrl.org

Remote Possibilities

The grand legacy of Amateur Radio is one of service and innovation. We can be proud of our history but we can scarcely afford to rest on our past accomplishments. In that spirit, our leadership launched major initiatives half a decade ago that are now bearing fruit and that still point to the future. Those initiatives — educational, technical and political alike — embody a good balance between what we communicate and how we communicate it. The social aspects tend to get more coverage than the technical, but we continually face and exercise unique opportunities to harmoniously combine them.

Four topics have been near the top of our agenda: digital voice, high-speed multimedia, software radio, and antenna restrictions. Each by itself is experiencing rapid and significant revolution. Combined, they represent a solution that doesn't have to wait for a problem: remote control.

We operators in antenna-restricted areas must fight a good fight on the legal front but we have the technology to circumvent the issue. The cost of awesome computing power in the form of personal computers has come within reach of nearly every ham. Our access to high-quality digital voice and image is now unprecedented, as largely driven by the cellular-phone and entertainment industries, as well as the Internet. High-speed data links are now routine, both over radio and on land lines. Software radio is just beginning to explore a realm of unlimited possibilities.

Put it all together and you have a solution to the antenna restriction problem and a whole lot more. Transmitters and receivers need not be colocated: They can be placed at optimal locations. Remote operation helps you avoid interfering with your neighbors' TVs, telephones and radios. It may let you dodge interference (BPL!) on receive, too. It allows clubs and individuals to pool their resources to get better equipment. With the proper gear, you can operate your remote rig from just about any location on Earth. Practical remote control of amateur rigs is a reality that's not getting enough traction.

We acknowledge that some knob-twisters don't like controlling their rigs with a computer, but there's no longer any reason to sacrifice functionality for remote control. To date, most of the practical systems we've seen are optimized for phone operation and little else. What we need is a universal system that supports not only voice operation, but also CW and three or four ancillary systems, such as an antenna rotator, a RTTY or other modem, a power amplifier and so on, perhaps using EIA-232 serial ports. The system must incorporate positive feedback and telemetry on the control link. It must comply with the Section 97.213(b) rule about a three-minute transmit limit should the control link fail. Finally, and perhaps most importantly, it ought to be capable of operation with any brand of transceiver having a standard digital control port using existing manufacturers' control software.

Except for certain software radios, manufacturers have largely ignored the call to publish application programming interfaces (APIs) that would standardize software control of their equipment and make updates a cinch. We feel, therefore, that what's needed is software that effectively extends all relevant PC functions, such as two channels of audio, serial and parallel ports between remote and control sites; and hardware for each site that facilitates interface. It would be nice if that system supported control links using Internet Protocol (IP), dial-up serial modems and VHF or above radios, including 802.11 types.

Yes, all of that can be and is being done now using a hodgepodge of off-the-shelf hardware and software, but we owe it to ourselves to get together and integrate things in a universal way. It's a neat chance to bring together those individual technologies that are proving so useful elsewhere.

Note that the recent FCC Report and Order in Docket 04-140, which went into effect December 15, 2006, allows auxiliary station operation (remote control links) on the 2-m band. Previously, auxiliary operation was permitted over links only on 1.25-m and shorter-wavelength bands. The Commission specifically cited enhanced possibilities for remote operation in its decision to modify Section 97.201(b) of the rules.

What do *you* think? Write us.

Octave for Transmission Lines

This installment in our series about the Octave *program looks at measuring feed line and antenna impedance.*

Maynard Wright, W6PAP

The ability to measure impedances is a wonderful thing. Various companies are making impedance bridges for the purpose. When combined with antenna design and analysis software, an impedance bridge can enable us to design and test antennas and transmission lines much more rapidly and conveniently than has been the case in the past.

The point at which we would like to measure impedances, though, is often out of reach of our instruments. The center insulator of a dipole 40 feet up is a difficult stretch for most of us. We must therefore measure antenna impedances through various lengths of transmission line. The problem with that is that the transmission line acts as an impedance transformer and we really won't be measuring the impedance at the antenna except in the case where the impedance is exactly matched at the antenna.

Does that matter? If all we care about is the match at the transmitter end of the line, maybe not. We can usually match that point pretty well, though, by using a tuner. The principal reasons for measuring antenna impedance is so that we can make sure that the line is reasonably well matched to the antenna to prevent a high SWR on the line and to make sure that the antenna is going to perform as we intended.

To refer a measurement at the near end of a transmission line to the distant end, we can use a Smith Chart as described on pages 21.4 and 21.5 of *The 2007 ARRL Handbook*.[1] The "Reflections on the Smith Chart" sidebar from *The ARRL Handbook* is included

[1]Notes appear on page 8.

6930 Enright Dr
Citrus Heights, CA 95621-2837
w6pap@arrl.net

in this article for the reader's convenience. Alternatively, we can use Eq 11 on page 21.9 of the *Handbook*:

$$Z_{in} = Z_0 \times \frac{Z_L \, cosh\,(\eta\,\ell) + Z_0 \, sinh\,(\eta\,\ell)}{Z_L \, sinh\,(\eta\,\ell) + Z_0 \, cosh\,(\eta\,\ell)}$$

(Eq 1)

where:

Z_{in} = the input impedance of the line

Z_L = the impedance seen at the terminating end of the line

Z_0 = the characteristic impedance of the line

η = the complex loss coefficient of the line (also known as the "complex propagation constant")

ℓ = the length of the line in the same units as those used to define η

As the *Handbook* points out, this equation is pretty tedious to apply by hand. Making serious transmission-line calculations back in the days of slide rules and trigonometry tables involved many hours of work for a few results. Today, we can ease the task using a programmable calculator or a computer. In fact, various folks have done that for us and ARRL's very capable *TLA* (Transmission Line Analysis) program is included on the CD-ROM that accompanies the *2007 Handbook*.

There is some advantage to rolling our own code, though. We can customize it as we wish. If we want to input the distance in furlongs, we can modify our code to handle that. In addition, we can learn something about the math involved in the problems we are solving if we write our own code, even if we make use of the software provided by ARRL or others for day-to-day calculations. The mathematical analysis tool *Octave*[2-5] allows us to do this without needing to know how to use a high-level language such as *C*.

The *Octave* code in Table 1 implements code for calculating the impedance at some specified point along a transmission line. A couple of housekeeping items are in order before you examine the code. Lines prefixed with the character "#" are comment lines and are ignored by *Octave* when a program is executed. The same is true of portions of a line of code beyond a "#" unless the "# is included inside a quoted string such as in a printf() statement. An ellipsis ("...") at the end of a line means that the next line is a continuation of the line with the ellipsis.

Note that only a few lines of code actually implement the equation we are interested in solving. The rest provide for input from the user and for formatted output of the results much as was the case in the previous *Octave* code we examined.

You might notice in Table 1 that there is another version of the input impedance equation that is commented out with a leading "#". It's much shorter than the one from the *Handbook* and you can use it if you like as the two expressions will yield identical answers. The commented out code is a convenient form (see Note 6, p 130), but it was not very useful in past years because early implementations of high-level languages like *FORTRAN* did not include inverse hyperbolic functions.[7] In addition, software implementations of inverse functions such as arctanh() usually return the principal value among many possibilities, and *Octave* is no exception in this regard. In some applications, this can lead to errors or uncertain results, but it shouldn't bother us here.

The code in Table 1 formats the input data for use by the input impedance equation. The equation wants the attenuation constant in nepers per unit length, while it's more convenient for us to input that characteristic of the line in decibels per unit length. We make

the conversion by multiplying by 0.1151 (see Note 6, pp 35-37). In addition, the code converts the velocity factor we specified to the phase constant of the line in radians per unit length. First, the wavelength in free space is calculated by dividing the speed of light in a vacuum by the frequency (see Note 6, p 31). The wavelength is then adjusted by multiply-

ing it by the velocity factor expressed as a fraction. Finally, the phase constant that the equation needs is calculated by dividing two pi by the wavelength (see Note 6, p 30).

Although the code could have asked for the input of the complex terminating impedance directly in one step, it's a bit more user friendly to ask separately for the real and

imaginary parts and to then combine them, as our *Octave* program does in the last line of code before calculation of the input impedance. The printf() statements that *Octave* uses to write the output in this program operate in a loop when more than one output number is present and pull the data from the proper row and column in the matrix that contains

---

Table 1

Octave **Code for Transmission Line Analysis**

```
# Print header

printf("\n\n     *** TRANSMISSION LINE CALCULATIONS ***\n");

# Enter input data from keyboard

f = input("\n                               FREQUENCY IN MHz: ");
d = input("                          LENGTH OF LINE IN FEET: ");
a = input("                       ATTENUATION IN dB PER 100 FEET: ");
v = input("                    VELOCITY FACTOR AS A PERCENTAGE: ");
Zo = input("                   CHARACTERISTIC IMPEDANCE IN OHMS: ");
Rt = input("       REAL PART OF TERMINATING IMPEDANCE IN OHMS: ");
Xt = input("IMAGINARY PART OF TERMINATING IMPEDANCE IN OHMS: ");

# Convert inputs as required

a = a ./ 1e2;  # convert dB per 100 feet to dB per foot
a = 0.1151 .* a;  # convert dB to nepers
c = 9.836e8;  # speed of light in feet per second
lambda = c ./ (1e6 .* f);  # wavelength of signal in vacuum
lambda = (v ./ 1e2) .* lambda; # adjust lambda for velocity
B = (2 .* pi) ./ lambda; # calculate Beta
Zt = Rt .+ j .* Xt;  # calculate complex terminating impedance

# Calculate input impedance

#Zd = Zo .* tanh((a .+ j .* B) .* d .+ atanh(Zt ./ Zo));

Zd = Zo .* ((Zt .* cosh((a .+ j .*B) .* d) .+ Zo .* ...
sinh((a .+ j .*B) .* d)) ./ (Zt .* sinh((a .+ j .*B) .* d) ...
.+ Zo .* cosh((a .+ j .*B) .* d)));

# Print results

for k = 1:columns(Zd)
   if imag(Zd(1,k)) < 0
      printf("\n\n  INPUT IMPEDANCE = %8.5g - j%-8.5g\n\n", ...
      real(Zd(1,k)), abs(imag(Zd(1,k))));
   else
      printf("\n\n  INPUT IMPEDANCE = %8.5g + j%-8.5g\n\n", ...
      real(Zd(1,k)), imag(Zd(1,k)));
   endif
endfor

# End input impedance program
```

---

the output data. The sign of the imaginary component is tested and, if negative, a separate printf format is used to make the output a little more presentable than would otherwise be the case.

One difference you will probably notice between this *Octave* code and the code listed in previous articles about *Octave* is that most of the arithmetic operators in Table 1 are preceded by a period ("."). That tells *Octave* that we want an element-by-element operation on the data rather than a matrix operation. *Octave* features some powerful matrix handling capabilities, but we want to hold them in check here and use only element-by-element operations. In matrix multiplication, two matrices may be multiplied if one has x rows and y columns and the other has y rows and x columns. If the two matrices do not meet those criteria, they may not be subjected to matrix multiplication.

To test this in *Octave*, invoke *Octave* from the command line and, at the prompt (octave:1>), type:

    x = [1,2,3,4]

Octave will respond with:

    x =
     1 2 3 4

Now type y = [5;6;7;8]

Octave will respond with:

    y =
     5
     6
     7
     8

Table 2

```
      *** TRANSMISSION LINE CALCULATIONS ***
                              FREQUENCY IN MHz:  7
                        LENGTH OF LINE IN FEET:  15
              ATTENUATION IN dB PER 100 FEET:    0
            VELOCITY FACTOR AS A PERCENTAGE:  66
            CHARACTERISTIC IMPEDANCE IN OHMS:  50
        REAL PART OF TERMINATING IMPEDANCE IN OHMS:  69.1
  IMAGINARY PART OF TERMINATING IMPEDANCE IN OHMS:  65.1

              INPUT IMPEDANCE =    40.24  -  j50.845
```

Table 3

```
      *** TRANSMISSION LINE CALCULATIONS ***
                              FREQUENCY IN MHz:  [7.01, 7.1, 7.2, 7.29]
                        LENGTH OF LINE IN FEET:  40
                ATTENUATION IN dB PER 100 FEET:  [0.75, 0.76, 0.76, 0.77]
              VELOCITY FACTOR AS A PERCENTAGE:  83
            CHARACTERISTIC IMPEDANCE IN OHMS:  75
        REAL PART OF TERMINATING IMPEDANCE IN OHMS:  [72.4,75,77.9,80.7]
  IMAGINARY PART OF TERMINATING IMPEDANCE IN OHMS:  [-0.46,21.3,44.4,65]

  The answer is given as:

      INPUT IMPEDANCE   =         76.345 - j2.1434
      INPUT IMPEDANCE   =         57.852 - j3.0632
      INPUT IMPEDANCE   =         44.027 + j1.6856
      INPUT IMPEDANCE   =         35.738 + j7.5798
```

Table 4

```
      *** TRANSMISSION LINE CALCULATIONS ***
                              FREQUENCY IN MHz:  [7.01, 7.1, 7.2, 7.29]
                        LENGTH OF LINE IN FEET:  -40
                ATTENUATION IN dB PER 100 FEET:  [0.75, 0.76, 0.76, 0.77]
              VELOCITY FACTOR AS A PERCENTAGE:  83
            CHARACTERISTIC IMPEDANCE IN OHMS:  75
        REAL PART OF TERMINATING IMPEDANCE IN OHMS:  [76.345,57.852,44.027,35.738]
  IMAGINARY PART OF TERMINATING IMPEDANCE IN OHMS:  [-2.1434,- 3.0632,1.6856,7.5798]

  The answer is given as:

      INPUT IMPEDANCE   =         72.4 - j0.46011
      INPUT IMPEDANCE   =         75 + j21.299
      INPUT IMPEDANCE   =         77.9 + j44.401
      INPUT IMPEDANCE   =         80.7 + j65
```

Reflections on the Smith Chart

Although most radio amateurs have seen the Smith Chart, it is often regarded with trepidation. It is supposed to be complicated and subtle. However, the chart is extremely useful in circuit analysis, especially when transmission lines are involved. The Smith Chart is not limited to transmission-line and antenna problems.

The basis for the chart is Eq 4 in the *Handbook* main text relating reflection coefficient to a terminating impedance. That equation is repeated here as Equation A.

$$\rho = \frac{Z - Z_0}{Z + Z_0}$$

(Eq A)

where Z_0 is the characteristic impedance of the chart, and $Z = R + j X$ is a complex terminating impedance. Z might be the feed-point impedance of an antenna connected to a Z_0 transmission line.

It is useful to define a normalized impedance $z = Z/Z_0$. The normalized resistance and reactance become $r = R/Z_0$ and $x = X/Z_0$. Inserting these into Eq 1 yields:

$$\rho = \frac{z - 1}{z + 1}$$

(Eq B)

where r and z are both complex, each having a magnitude and a phase when expressed in polar coordinates, or a real and an imaginary part in XY coordinates.

Equations A and B have some interesting and useful properties, characteristics that make them

Vel. Fac. = 0.6600
Z0 = 50.0000

Fig B—This plot shows a Smith Chart. The circles now represent contours of constant normalized resistance. The arcs represent constant normalized reactance. Note the arc with the markers: This illustrates the same antenna and line used in the previous figure. The plot is the same on the two charts; only the scale details have changed.

QX0701-WrightB

Fig A—Plot of polar reflection coefficient. Circles represent contours of constant ρ. The starting "feed point" value, 0.5 at +45°, represents an antenna impedance of 69.1 + *j* 65.1 Ω with Z_0 = 50 Ω. The arc represents a 15-ft section of 50-Ω, VF 0.66 transmission line at 7 MHz, yielding a shack ρ of 0.5 at −71.3°. The shack z is calculated as 40.3 − *j* 50.9 Ω.

QX0701-WrightC

Fig C—The Smith Chart shown in Fig B was computer generated. A much more detailed plot is presented here; this is the chart form used by Smith, suitable for graphic applications. This chart is used with the permission of Analog Instruments.

physically significant:

- Even though the components of z (and Z) may take on values that are very large, the reflection coefficient ρ, is restricted to always having a magnitude between zero and one if z has a real part, r, that is positive.
- If all possible values for ρ are examined and plotted in polar coordinates, they will lie within a circle with a radius of one. This is termed *the unit circle*. A plot is shown in **Fig A**.
- An impedance that is perfectly matched to Z_0, the characteristic value for the chart, will produce a ρ at the center of the unit circle.
- Real Z values, ones that have no reactance, "map" onto a horizontal line that divides the top from the bottom of the unit circle. By convention, a polar variable with an angle of zero is on the x axis, to the right of the origin.
- Impedances with a reactive part produce ρ values away from the dividing line. Inductive impedances with the imaginary part greater than zero appear in the upper half of the chart, while capacitive impedances appear in the lower half.
- Perhaps the most interesting and exciting property of the reflection coefficient is the way it describes the impedance-transforming properties of a transmission line, presented in closed mathematical form in the main text as Eq 1. Neglecting loss effects, a transmission line of electrical length θ will transform a normalized impedance represented by ρ to another with the same magnitude and a new angle that differs from the original by -2θ. This rotation is clockwise.

Clearly, the reflection coefficient is more than an intermediate step in a mathematical development. It is a useful, alternative description of complex impedance. However, our interest is still focused on impedance; we want to know, for example, what the final z is after transformation with a transmission line. This is the problem that

Phillip Smith solved in creating the Smith Chart. Smith observed that the unit circle, a graph of reflection coefficient, could be labeled with lines representing *normalized impedance*. A Smith Chart is shown in **Fig B**. All of the lines on the chart are complete or partial circles representing a line of constant normalized resistance and reactance.

How might we use the Smith Chart? A classic application relates antenna feed-point impedance to the impedance seen at the end of the "shack" end of the line. Assume that the antenna impedance is known, $Z_a = R_a + j X_a$. This complex value is converted to normalized impedance by dividing R_a and X_a by Z_0 to yield $r_a + j x_a$, and is plotted on the chart. A compass is then used to draw an arc of a circle centered at the origin of the chart. The arc starts at the normalized antenna impedance and proceeds in a clockwise direction for $2\theta°$, where θ is the electrical degrees, derived from the physical length and velocity factor of the transmission line. The end of the arc represents the normalized impedance at the end of the line in the shack; it is denormalized by multiplying the real and imaginary parts by Z_0.

Antenna feed-point Z can also be inferred from an impedance measurement at the shack end of the line. A similar procedure is followed. The only difference is that rotation is now in a counterclockwise direction. The Smith Chart is much more powerful than depicted in this brief summary. A detailed treatment is given by Phillip H. Smith in his classic book: *Electronic Applications of the Smith Chart* (McGraw-Hill, 1969). I also recommend his article "Transmission Line Calculator" in Jan 1939 *Electronics*. Joseph White presented a wonderful summary of the chart in a short but outstanding paper: "The Smith Chart: An Endangered Species?" Nov 1979 *Microwave Journal*.
—*Wes Hayward, W7ZOI*

You have now created two matrices: *x* with one row and four columns and *y* with four rows and one column. We can use matrix multiplication to multiply them:

z = x * y
z = 70

This is the correct result of matrix multiplication of the two matrices. If we try to multiply them element-by-element using ".*" we will draw an error message because the two matrices must have the same number of rows and columns to be multiplied element by element.

Now reassign y so that it has the same format as x:

y = [5, 6, 7, 8]
y =
 5 6 7 8

Now multiply them element-by-element:

z = x .* y

The result is:

z =
 5 12 21 32

Notice that the first two elements of each matrix were multiplied to produce the first element of the result and so forth. We will use this element-by-element feature in making our transmission line calculations.

Whether or not you include the period before the operator, any matrix may be multiplied by a constant. For this reason, we could have omitted the periods from some of the operators in the code in Table 1, but it's easier to include them everywhere in cases where they don't cause any trouble and where omitting one in the wrong place could cause errors.

Before we proceed, let's check our code. We'll use the Smith Chart example in the caption for Figure A in the "Reflections on the Smith Chart" sidebar. If you've entered the code from Table 1 into a file in your computer named "trs_line.m," invoke the program by typing "octave trs_line.m" and press the ENTER key. You should see:

*** TRANSMISSION LINE CALCULATIONS ***

FREQUENCY IN MHz:

Since the text in Figure A specifies 7 MHz, enter 7 and press the ENTER key. Table 2 shows the display, including the results after you have entered all of the information from Figure A in the sidebar.

That's pretty close to the $40.3 - j50.9$ yielded by the Smith Chart example from the *Handbook* and we'll accept it as evidence that our code is not going astray.

Now let's look at a typical antenna. We'll generate a 40-meter dipole using *nec2c*.[8] We get the following impedances at the antenna:

Freq in MHz	Impedance at Center
7.010	$72.4 - j0.46\ \Omega$
7.100	$75.0 + j21.3$
7.200	$77.9 + j44.4$
7.290	$80.7 + j65.0$

If we were attempting to create a resonant antenna for the low end of 40, we did pretty well. Note, though, that we will be sacrificing some of the usable SWR range of our antenna by letting it fall below the lower limit of the 40-meter band.

Now we need to decide on a transmission line. Let's use 40 feet of Belden 1426A. We'll obtain the characteristics of that line from Table 21.1 of the *2007 Handbook*:

Characteristic Impedance: 75 Ω
Velocity Factor: 83%

The other characteristic we need is the attenuation in dB per 100 feet, which is specified at several different frequencies, but not at our 40-meter frequencies. The attenuation of most transmission lines is approximately proportional to the square root of frequency at high frequencies. Rather than curve-fitting the data in the *Handbook* precisely, we can assume that an equation for the attenuation of our line will have the following form:

$atten = k * sqrt(f)$

where *atten* is the attenuation in dB per 100 feet, *f* is the frequency in MHz, and *k* is a constant. We rearrange the equation to solve for *k*:

$k = atten / sqrt(f)$

We solve for *k* at 10 MHz and obtain *k* = 0.285. Plugging this back into the original equation at the other frequencies specified in the *Handbook*, we find that we are pretty close, so we'll use this equation. Applying it at our 40-meter frequencies, we get:

Freq in MHz	Attenuation in dB per 100 feet
7.010	0.75
7.100	0.76
7.200	0.76
7.290	0.77

Now we're ready to calculate the input impedance of our transmission line at the frequencies at which we've specified our antenna. We do that by running the *Octave* code and entering data as shown in Table 3.

The rows with the input impedances are listed in the order of the columns of the input. If we were to add a calculation at 6.99 MHz here, we would see that the imaginary component of input impedance has started to rise and has become positive. There is a minimum at about 7.1 MHz. Note that the fact that the imaginary component of the antenna impedance crosses zero at only one point doesn't force the same result onto the impedance some distance away along a transmission line.

Well, this seems to work fine, but what good is it? We started by assuming an array of antenna impedances and, in the real world, we don't have that data. Just as with the Smith Chart, though, we can use the *Octave* program to move either direction along a transmission line. Let's assume that we don't know the antenna impedance, but that we have measured the data we just calculated. If we use those numbers as input data and move in the opposite direction by specifying a negative length, we should get the impedances at the antenna. See Table 4.

These numbers are, neglecting roundoff error, the antenna impedances we submitted to *Octave* for the original calculation.

If we were to map our calculations onto a Smith Chart, we would see a spiral path traversing somewhat over one quarter wavelength around the Chart. As we proceed from the antenna toward the analyzer end of the line, the trace would move toward the center of the Chart due to the attenuation of the line. In our case, the effect of the attenuation is not very pronounced, but it does make a difference. If we recompute the input impedance at 7.01 MHz, but with zero attenuation and

the same antenna impedance we used before, we get 76.4 + *j*2.109 Ω.

Since we are working with SWR values pretty close to 1:1 here, the action on the Smith Chart will consist of small spirals near the center of the chart. My Smith Chart slide rule (Reference 9) won't allow me to work that close to the center of the Chart, so I would need to use a paper Smith Chart to replicate the calculations we have done here.

By changing the mathematics a little, we can produce outputs in terms of SWR or reflection coefficient if we like. We might also revise the input requests to accept transmission line data in terms of primary or secondary constants if we have transmission line data available in either of those formats.

We haven't done anything here that we can't do with any of several more powerful and convenient programs that are already available to us. We've learned, though, a little bit about how to handle transmission line calculations and we've learned that *Octave* includes some powerful facilities for handling math calculations and for manipulating matrices of data.

Notes

[1]*The 2007 ARRL Handbook for Radio Communications*, The American Radio Relay League, Inc, 2006.

[2]M. Wright, W6PAP, *Octave — Calculations for Amateurs*, QEX, May/Jun 2005, pp 48-50.

[3]M. Wright, W6PAP, *Octave for Signal Analysis*, QEX, Jul/Aug 2005, pp 57-61.

[4]M. Wright, W6PAP, *Octave for System Modeling*, QEX, Jul/Aug 2006.

[5]John W. Eaton, *GNU Octave Manual, Network Theory Limited*, 1997 (see **www.octave.org**).

[6]Robert A. Chipman, *Schaum's Outline Series, Theory and Problems of Transmission Lines*, McGraw-Hill, 1968.

[7]Daniel E. Alexander and Andrew C. Messer, *Fortran IV Pocket Handbook*, McGraw-Hill, 1972.

[8]*nec2c* is a port of *NEC2* from *FORTRAN* to *C* by Ray Anderson, WB6TPU. *nec2c* may be downloaded from **www.si-list.org/swindex2.html**.

[9]My Smith Chart slide rule was manufactured years ago by the Emeloid Company of New Jersey. I don't know whether or not such slide rules are still available.

Maynard Wright, W6PAP, was first licensed in 1957 as WN6PAP. He holds an FCC General Radiotelephone Operator's License with Ship Radar Endorsement, is a Registered Professional Electrical Engineer in California, and is a Senior Member of IEEE. He has been involved in the telecommunications industry for over 43 years. He has served as technical editor of several telecommunications standards, and holds several patents. He is a Past Chairman of the Sacramento Section of the IEEE. Maynard is a member of the North Hills Radio Club in Sacramento, California, where he currently serves as Secretary/Treasurer, Net Manager and is a Past President. **QEX**

An Alternative Transmission Line Equation

The use of exponential or hyperbolic functions in the transmission line equation is a matter of choice rather than necessity. An alternative equation that uses neither is presented here.

Ron Barker, G4JNH, VK3INH

The term transmission line equation can be used to describe any equation which relates input impedance, load impedance, characteristic impedance, electrical length and loss. Chipman, in his *Theory and Problems of Transmission Lines*, lists no less than five such equations all of which deploy either exponential or hyperbolic functions.[1] The version which appears to have been most widely used is that which Dr Steven R. Best, VE9SRB, described in the first of his series of three excellent articles which appeared in *QEX* in 2001.[2] This is how he presented it:

$$Z_{IN} = Z_0 \left[\frac{\dfrac{Z_A}{Z_0} + \tanh(\gamma L)}{1 + \dfrac{Z_A}{Z_0}\tanh(\gamma L)} \right] \qquad \text{(Eq 1)}$$

where:
Z_0 = characteristic impedance (Ω).
Z_A = load impedance (Ω).
Z_{IN} = input impedance (Ω).
L = line length in meters.
γ = line propagation factor ($\alpha + j\beta$).
and:
α = line loss in nepers per meter.
β = angular electrical length of line in radians per meter.

The potential usefulness of this equation in applications such as the design of transmission line matching sections or in remote impedance measurement is self-evident but its application involves math which will be unfamiliar to anyone whose math education stopped below university level. The problem arises from the need to establish the value of $\tanh(\gamma L)$ because (γL) is a complex number and ordinary scientific calculators, even including those which will handle

[1]Notes appear on page 14.

171 Leicester Rd
Ashby de la Zouch
Leicestershire LE65 1TR
Great Britain
g4jnh@onetel.com

complex number arithmetic, give the error sign when asked to give the hyperbolic function of imaginary or complex numbers. Even the mighty Microsoft *Excel* is not set up to do it. So unless you have a very special calculator or sophisticated math software, you have to know how to expand the hyperbolic function into components that an ordinary scientific calculator can handle. (See Appendix A.)

There is another option. You could just forget the equation and key your numbers into one of the many software programs available that would do it for you. I have no problem with using such software but I find it frustrating when I don't understand the underlying math. At the time that I read Dr Best's article, I could handle the complex arithmetic but did not know how to handle the complex hyperbolic function and couldn't find anything on my bookshelves that covered it. Furthermore, I was keen to be able to understand Dr Best's derivation of the equation, particularly as his subheading described it as "simple and elegant." I therefore resolved to learn the necessary math. Naively, I didn't envisage any problem since amongst family, friends, acquaintances and ex-colleagues there were at least half a dozen who had read math at university level. But I was to be disappointed as I got the same unhelpful answer from them all: "I must have known it at the time but I can't remember it now."

However, one of them, my son, unearthed a university math textbook that I now have on permanent loan.[3] This book was written for students who were being tutored, rather than as a teach yourself manual and I found it very heavy going. However, it covered all of the relevant subject matter and with considerable effort I was eventually able to make sense of it all.

Having gotten to the stage of understanding the math of the derivation of the transmission line equation, I realized that the use of exponential or hyperbolic functions is a matter of choice rather than of necessity. The purpose of this article is to present an alternative transmission line equation which uses neither.

To understand any manipulation of complex impedance, it is essential to understand the basics of complex algebra, so the first part of the article is devoted to that. The second part of the article looks in detail at the voltage reflection coefficient and how the way it changes along the line governs the impedance changes along the line. The final part describes the derivation of an alternative transmission line equation from the reflection coefficients at the two ends.

Basics of Complex Algebra

Picture this scenario: You are driving along the freeway from your home to your place of work at 60 mph. The distance between you and your place of work is changing at a rate of –60 mph and the distance between you and your home is changing at a rate of +60 mph. If you change your mind and do a smart 180° U-turn without changing your speed (not recommended), the distance between you and your place of work is now changing at a rate of +60 mph and the distance between you and your home is changing at –60 mph. The important point to note here is that turning through 180° had the same effect on the sign of your speed as multiplying it by –1. Let us take another example involving the domestic electricity supply.

Here in the UK, the supply is nominally at 240 V_{rms} at 50 Hz, which means that at the positive peak of the cycle the instantaneous line voltage is $240 \times \sqrt{2} = + 339.4$ V. One hundredth of a second or half a cycle (180°) later it is –339.4 volts. Another example of 180° of rotation being equivalent to multiplying by –1. This simple and obvious relationship is the starting point for the branch of math unfortunately known as complex algebra — unfortunate because complex implies difficult, which it isn't.

It is opportune at this point to take our first look at the Argand diagram (Figure 1), named after its originator, the French mathematician Jean-Robert Argand (1768-1822). It looks like the coordinates of a perfectly ordinary x-y plot; but as we shall see, it is used rather differently. The scales of the x and y coordinates must be the same and the intersection of the coordinates must be at 0 for both. In Figure 1, both coordinates are scaled –1 to +1 but they could be anything provided that they are scaled the same. If we stick the point of a pair of drawing compasses in the common zero and the pencil at +1 on the x coordinate and then swing the compasses anticlockwise through 180° the pencil very obviously crosses the x coordinate on the negative side at –1. This is exactly the same relationship between turning through 180° and multiplying by –1 as we saw previously. Note that in making the 180° swing, the pencil crossed the y coordinate at +1 on a swing of 90°. What would we have to multiply by to have the same effect as turning through 90°? Whatever the number is, we know that if we did the multiplication twice, which would be equivalent to turning through 180°, it would be the same as multiplying by –1. This means that the square of the number we are looking for must be equal to –1, which in consequence, means that it must be equal to the square root of –1. And that presents a problem because, as we all learned at school, negative numbers don't have real square roots.

The first mathematicians to address the problem, albeit in a dif-ferent context, decided to imagine that –1 had a square root and not surprisingly designated it by the letter *i* (lower case) to mean *imaginary*. When electrical engineers started using complex algebra, they decided that, because I (upper case) had already become universally accepted as the symbol for current, it would avoid confusion if the letter *j* were used in electrical calculations. Mathematicians and mechanical engineers generally still use the *i* symbol, however. We shall use only *j* in this article.

All numbers that fall on the x axis of the Argand diagram are known as real numbers and all numbers falling on the y axis are prefixed by *j* and are known as imaginary numbers. It will be evident therefore that:

turning through 90° is equivalent to multiplying by *j*

turning through 180° is equivalent to multiplying by $j^2 = –1$

turning through 270° is equivalent to multiplying by $j^3 = j^2 \times j = –1 \times j = –j$

turning through 360° is equivalent to multiplying by $j^4 = j^2 \times j^2 = –1 \times –1 = 1$

We now need to consider what happens when we swing through angles which are not multiples of 90°. Turning again to the Argand diagram (Figure 2), let us take a look at what would be the result of starting at +1 on the x axis and swinging the drawing compasses through 45° to point A. Note that point A is a vector quantity, as are all points on the Argand diagram and its position can be defined in two ways. One way would be to specify it to be at a distance of 1 unit along a line originating at the common zero of the x and y axes, which is at an angle of 45° to the x axis. This is defined as the polar representation of the point. Alternatively, its position could be speci-fied in terms of the x and y coordinates as x + *j*y. This is known as the rectangular or scalar representation and x + *j*y is a complex number. The length of the line OA is known to mathematicians as the modulus or absolute of the complex number and is conventionally designated by the letter r. When used in electrical engineering it is referred to as the magnitude of the vector quantity. The angle between the line OA and the positive x axis (45° in this case) is known to mathematicians as the argument of the complex number and is conventionally designated by the Greek letter θ. In electrical engineering, it is referred to as the vector angle. It is deemed to be positive when the imaginary is posi-tive and negative when the imaginary is negative. It is equally valid to express all the angles as positive by adding 360° to the negative values. Throughout this article we shall use magnitude and angle to define vector quantities.

To convert from polar to rectangular representation we apply the rules of simple trigonometry as follows:
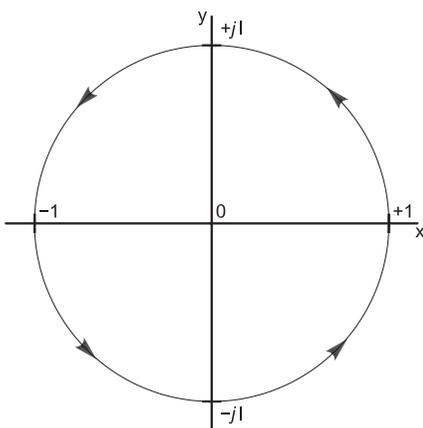
Figure 1 — Argand Diagram, see text.

QX0701-Bark01
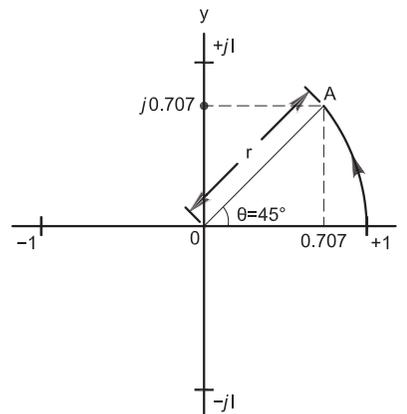
Figure 2 — Argand Diagram — Relationship between polar and rectangular representation of a complex quantity.

QX0701-Bark02

$x = r \cos\theta$ and:

$y = r \sin\theta$ so:

$$x + jy = r \cos\theta + jr \sin\theta \qquad \text{(Eq 2)}$$

To convert from rectangular to polar representation we use Pythagoras' theorem to determine the magnitude:

$$r = (x^2 + y^2)^{1/2} \qquad \text{(Eq 3)}$$

There are three equations for deriving the angle depending on whether the real is positive, negative or zero:

$$\theta = \tan^{-1}(y/x), \; x > 0 \qquad \text{(Eq 4A)}$$
$$\theta = \tan^{-1}(y/x) + 180°, \; x < 0 \qquad \text{(Eq 4B)}$$
$$\theta = 90°, \; x = 0 \qquad \text{(Eq 4C)}$$

With our values for point A of r = 1 and θ = 45°, Equation 2 gives:

$$x + jy = 0.707 + j0.707$$

which just happens to be the square root of j but that needn't concern us.

With the rectangular representation, both the real and imaginary components can be handled in the same mathematical expression as a complex number. To handle the components of the polar representation in the same mathematical expression, it is necessary to express them as an exponential function but we do not need to pursue that for the purpose of this article.

Note that the j operator is treated in the same way as any algebraic symbol except that when j^2 arises we change it to –1. To demonstrate this, we will add, subtract, multiply and divide two complex numbers, $2 + j3$ and $4 - j5$.

Addition: $(2 + j3) + (4 - j5) = 6 - j2$.

Subtraction: $(2 + j3) - (4 - j5) = -2 + j8$

Multiplication: $(2 + j3) \times (4 - j5) = 8 - j10 + j12 - j^2 15$
$= 23 + j2$.

Division: To accomplish division, it is necessary to eliminate the imaginary component from the denominator and this is achieved by multiplying both numerator and denominator by the complex conjugate of the denominator. The conjugate of a complex number has the same numerical values but the sign of the imaginary component is changed. So,

$$\frac{2 + j3}{4 - j5} = \left(\frac{2 + j3}{4 - j5}\right)\left(\frac{4 + j5}{4 + j5}\right)$$

$$= \frac{8 + j10 + j12 + j^2 15}{16 + j20 - j20 - j^2 25}$$

$$= \frac{8 + j22 - 15}{16 + 25}$$

$$= \frac{-7 + j22}{41}$$

$$= -0.171 + j0.537$$

Having looked at the simple arithmetic of complex numbers, we can now turn our attention to examining what happens to their polar representations when they are added, subtracted, multiplied and divided. Table 1 lists the numbers we have just worked with and their polar equivalents.

It can be readily shown, either by trigonometry or graphical means, that when two complex numbers are added, the polar representation of their sum is equal to the vector sum of their polar values. Similarly it can be shown that when two complex numbers are subtracted, the polar

representation of their difference is equal to the vector difference of their polar values. This is exactly as would be expected and you may like to prove it to yourself by drawing them out on a sheet of graph paper.

However, the situation becomes much more interesting when complex numbers are multiplied or divided. Looking first at multiplication, we see that in the polar representation the magnitudes are multiplied and the angles are added. In our example $3.605 \times 6.403 = 23.087$ and $56.30 + (-51.30) = 5.00$. Similarly, for division we see that in the polar representation the magnitudes are divided and the angles subtracted. In our example $3.605/6.403 = 0.564$ and $56.3 - (-51.3) = 107.6$.

But — and this is very important in the context of the alternative transmission line equation — if we wish to multiply the magnitudes and subtract the angles we can achieve this in rectangular form by reversing the sign of the imaginary part of the multiplier. Similarly, if we wish to divide the magnitudes and add the angles we can also achieve this in rectangular form by reversing the sign of the imaginary part of the divisor.

That is all that we need to know about complex numbers to understand the rest of this article and we can now proceed to a detailed study of the voltage reflection coefficient.

Voltage Reflection Coefficient

Voltage reflection coefficient is defined as being the ratio of the reflected voltage to the forward voltage on a transmission line terminated by an unmatched load:

$$\rho = \frac{V_R}{V_F} \qquad \text{(Eq 5)}$$

where:

ρ = voltage reflection coefficient

V_R = reflected voltage

V_F = forward voltage

Although all of these are vector quantities (indicated by boldface type) having both magnitude and angle, the term reflection coefficient is frequently taken to mean its magnitude which is more correctly defined as follows:

$$|\rho| = \frac{|\mathbf{V_R}|}{|\mathbf{V_F}|} \qquad \text{(Eq 5A)}$$

However, in the study of transmission line behavior, both the magnitude and angle are of equal importance. We will now see how these are related to the impedance at an unmatched termination by using Thevenin's Theorem. Put in simple terms, a Thevenin equivalent circuit consists of a constant voltage source, V_S, in series with a specified output impedance, which in this case is the characteristic impedance of the line, Z_0. Because a transmission line will only propagate electrical energy at a specific ratio of voltage to current as defined by its characteristic impedance, a Thevenin equivalent circuit analysis is valid. The fact that the characteristic impedance is merely

Table 1

Rectangular representation Polar representation

Complex Impedance	Absolute (Magnitude)	Argument (angle)
$2 + j3$	3.605	56.30°
$4 - j5$	6.403	–51.30°
Sum = $6 - j2$	6.324	–18.40°
Difference = $-2 + j8$	8.246	–76.00°
Product = $23 + j2$	23.083	5.000°
Quotient = $-0.171 + j0.537$	0.563	107.6°

a ratio of voltage to current rather than a dissipative resistance does not invalidate this circuit analysis.

The Thevenin equivalent circuit of a transmission line is shown in Figure 3. By application of Ohm's Law:

$$\mathbf{V_A} = \mathbf{V_S} \left(\frac{\mathbf{Z_A}}{\mathbf{Z_A} + \mathbf{Z_0}} \right) \tag{Eq 6}$$

where V_A = voltage across Z_A.

It will be evident, therefore, that when the line is terminated in a matched load, the voltage across the load will be the same as the forward voltage and will be equal to half the source voltage:

$$V_F = V_{A(\text{matched load})} = V_S / 2 \tag{Eq 7}$$

But we know that the voltage across the load is equal to the sum of the forward and reflected voltages:

$V_A = V_F + V_R$ so:
$$V_R = V_A - V_F \tag{Eq 8}$$

Substituting Equation 8 into Equation 5 for the reflection coefficient at the load gives:

$$\rho_A = \frac{V_A - V_F}{V_F} \tag{Eq 9}$$

Substituting Equations 6 and 7 into Equation 9 gives:

$$\rho_A = \frac{V_S \left(\dfrac{Z_A}{Z_A + Z_0} \right) - \dfrac{V_S}{2}}{\left(\dfrac{V_S}{2} \right)}$$

$$= \frac{2Z_A}{Z_A + Z_0} - 1$$

$$= \frac{2Z_A}{Z_A + Z_0} - \frac{Z_A + Z_0}{Z_A + Z_0}$$

$$= \frac{Z_A - Z_0}{Z_A + Z_0} \tag{Eq 10}$$

This is the familiar equation for complex reflection coefficient.

Let us now apply this equation to a typical situation with RG-213 transmission line. The details are shown in Table 2.

Putting the above values into Equation 1 shows the input impedance to be $30.6 + j10.0 \ \Omega$. An explanation of how this result is worked out is presented in Appendix B. If we enter the above Z_A and Z_0 values into Equation 10, we get: $\rho_A = -0.08266 - j0.3152$. We can now determine the magnitude and angle using Equations 3 and 4B (the real is negative) as follows:

$$\left| \rho_A \right| = \left(0.08266^2 + 0.3152^2 \right)^{\frac{1}{2}} \tag{Eq 11A}$$
$$\left| \rho_A \right| = 0.3259$$

$$\theta_A = \tan^{-1} \left(\frac{-0.3152}{-0.08266} \right) + 180° \tag{Eq 11B}$$
$$\theta_A = 255.3°$$

The next step is to enter this result on an Argand diagram as shown in Figure 4.

We now need to examine what happens to the reflection coefficient vector as we move back along the line towards the input. As we move back along the line, the forward and reflected voltages rotate in opposite directions relative to each other so the angle between them, which is

Table 2

Characteristics of Sample of RG-213

Characteristic impedance, Z_0	$50.0 - j0.3 \ \Omega$
Length	30.0 meters
Matched loss	0.8 dB
Velocity factor	0.66
Frequency	14.2 MHz
Load impedance, Z_A	$35.0 - j25.0 \ \Omega$
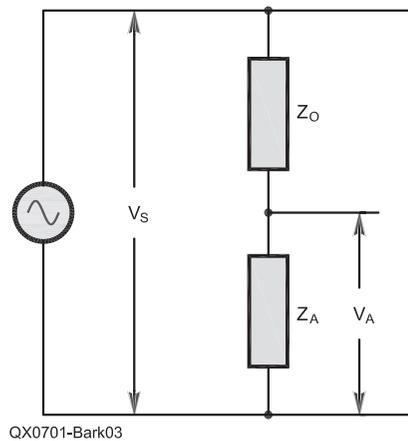
Figure 3 — Thevenin equivalent circuit.

QX0701-Bark03

QX0701-Bark04
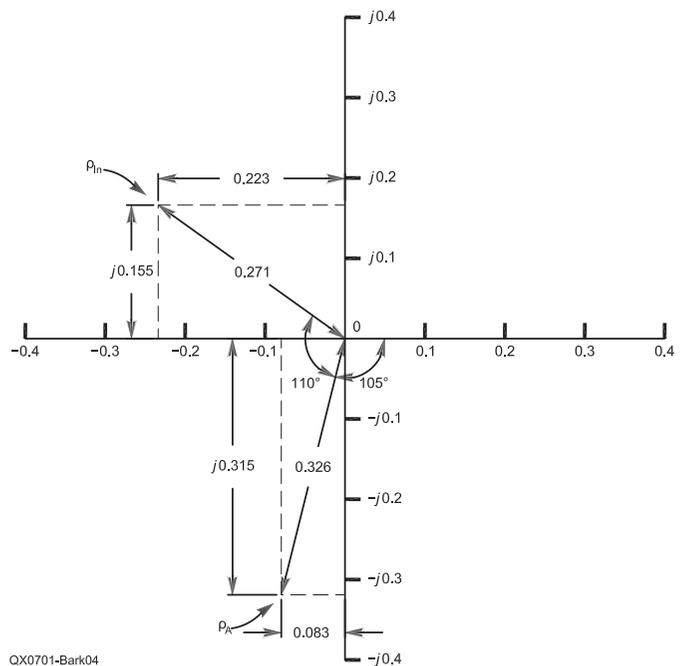
Figure 4 — Argand Diagram — showing change in reflection coefficient between input and load.

the angle of reflection coefficient, rotates at twice the angular length of the line. In other words, if we move back along the line by n electrical degrees, the angle of reflection coefficient rotates by $2n$ degrees. At the same time, the effect of line loss is to increase the forward voltage and reduce the reflected voltage by the matched line loss value. So at the input to the line, the difference between the magnitudes of the forward and reflected voltages is increased by twice the matched line loss, as expressed in dB; or the square of the matched line loss if it is expressed as a voltage multiplier. It follows, therefore, that we need to know the length of our RG-213 line expressed in electrical degrees and its loss expressed as a voltage multiplier.

Let:

a = loss as a voltage multiplier

b = angular electrical length in degrees

Then:

$$a = 10^{\frac{\text{loss in dB}}{20}} \qquad \text{(Eq 12)}$$

$$= 10^{\frac{0.8}{20}} = 0.9120$$

and:

$$b = 360° \left(\frac{\text{length} \times \text{frequency}}{\text{speed of light} \times \text{velocity factor}} \right) \qquad \text{(Eq 13)}$$

$$= 360° \left(\frac{30.0\,\text{m} \times 14.2\,\text{MHz}}{299.8\,\text{Mm/s} \times 0.66} \right) = 775.1°$$

As we saw earlier, the reflection coefficient vector rotates at twice the angular length of the line so we have to double it to obtain the angle by which it changes between the load and input ends of the line. That gives us a value of 1550.2°. To enter this on to the Argand diagram, we need to express it modulo-360: That comes out at 110.2. Because we are moving back along the line towards the input, we go round the Argand diagram in a clockwise direction, which means that the angle has to be subtracted from the angle of the reflection coefficient at the load end:

$\theta_{IN} = 255.3° - 110.2° = 145.1°$

The magnitude of the reflection coefficient at the input, $|\rho_{in}|$, is equal to $|\rho_A|$ multiplied by the square of the one-way line loss:

$|\rho_{in}| = 0.3259 \times 0.9120 \times 0.9120 = 0.2711$

This vector can now be added to Figure 4 and if you are a user of the Smith Chart, what we have just done will have a ring of familiarity, the reason being that the Smith Chart is a modified Argand diagram.[4]

Knowing the reflection coefficient at the input to the line enables us to determine the input impedance, but we must first convert it from the polar form to rectangular form using Equation 2:

$\rho_{IN} = |\rho_{IN}|\,cos\,\theta_{IN} + j|\rho_{IN}|\,sin\,\theta_{IN}$

$= 0.2711\,cos\,145.1° + j\,0.2711\,sin\,145.1°$

$= -0.223 + j0.155$

We can now return to Equation 10 to calculate Z_{IN}:

$$\rho_{IN} = \frac{Z_{IN} - Z_0}{Z_{IN} + Z_0}$$

$$\therefore Z_{IN} = Z_0 \left(\frac{1 + \rho_{IN}}{1 - \rho_{IN}} \right) \qquad \text{(Eq 14)}$$

Entering our values, we get $Z_{IN} = 30.6 + j10.0$, which is the same

result as was derived from Equation 1.

Let us summarize what we have just done. We took the angle of reflection coefficient at the load and subtracted from it an angle equal to twice the polar electrical length of the line to arrive at the angle of reflection coefficient at the input. We took the magnitude of reflection coefficient at the load and multiplied it by the square of the line loss to arrive at the magnitude of reflection coefficient at the input. We converted those values to rectangular form to give us the complex reflection coefficient at the input from which we were able to derive the input impedance.

Note that we used two transmission line parameters — an angle equal to twice its angular electrical length and a magnitude equal to the square of its loss as a voltage multiplier. This makes it a vector quantity, which means that it can be converted to rectangular form as a complex number by Equation 2. We also saw that when two complex numbers are multiplied, the result in polar form is the same as adding the angles and multiplying the magnitudes; but if the sign of the imaginary part of the multiplier is reversed, the result is the same as multiplying the magnitudes and subtracting the angles, which is what we require. So if we take Equation 2 and change the sign of the imaginary component to negative, it will give us a complex number which is a constant for the transmission line at a specified frequency. I suggest that for the purpose of this article we call it the impedance transformation constant and identify it by the letter k.

So:

$$k = a^2 \cos(2b) - ja^2 \sin(2b) \qquad \text{(Eq 16)}$$

where: a = line loss as a voltage multiplier (from Equation 12), and b = angular line length (from Equation 13).

If we insert the values for our 30-meter length of RG-213, we get:

$k = 0.912^2 \times \cos(775.06 \times 2) - j0.912^2 \times \sin(775.06 \times 2)$

$= 0.8317 \times -0.3436 - j0.8317 \times 0.9391$

$= -0.2858 - j0.7811$

The variable k serves the same purpose as γL in the hyperbolic transmission line equation (Equation 1) and has the following relationship to it that needn't concern us:

$$k = e^{-2\gamma L} \qquad \text{(Eq 17)}$$

If we multiply the complex reflection coefficient at the load by the impedance transformation constant, we should get the complex reflection coefficient at the input. Let's see what happens:

$\rho_{IN} = k\rho_A$

$= (-0.08266 - j0.3152)(-0.2858 - j0.7811) \qquad \text{(Eq 18)}$

$= -0.2226 + j0.1547$

This is the same result as we saw earlier. Equation 18 is the skeleton equation which we will now build on to derive the alternative transmission line equation.

The Alternative Transmission Line Equation

To derive the alternative transmission line equation, we will start with the skeleton equation (Equation 18) and substitute ρ_{IN} and ρ_A by the impedance expressions from which they were derived using Equation 10.

From Equation 10:

$$\rho_{IN} = \frac{Z_{IN} - Z_0}{Z_{IN} + Z_0}$$

and

$$\rho_A = \frac{Z_A - Z_0}{Z_A + Z_0}$$

Substituting into Equation 18 gives:

$$\frac{Z_{IN} - Z_0}{Z_{IN} + Z_0} = k\left(\frac{Z_A - Z_0}{Z_A + Z_0}\right)$$

$$\therefore \quad (Z_{IN} - Z_0)(Z_A + Z_0) = k(Z_{IN} + Z_0)(Z_A - Z_0)$$

$$Z_{IN}Z_A + Z_{IN}Z_0 - Z_AZ_0 - Z_0^2 = kZ_{IN}Z_A - kZ_{IN}Z_0 + kZ_AZ_0 - kZ_0^2$$

$$Z_{IN}(Z_A + Z_0 - kZ_A + kZ_0) = Z_0(kZ_A - kZ_0 + Z_A + Z_0)$$

$$Z_{IN} = Z_0\left[\frac{Z_A + Z_0 + k(Z_A - Z_0)}{Z_A + Z_0 - k(Z_A - Z_0)}\right] \qquad \text{(Eq 19)}$$

So that's the alternative transmission line equation. Now we need to test it using the parameters of our length of RG-213 which were $Z_0 = 50.0 - j0.3$, $Z_A = 35.0 - j25.0$, $k = -0.2858 - j0.7811$ (from Equation 16): $Z_{IN} = 30.6 + j10.0$ ohms, which is the result that we were expecting.

So far we have only considered the situation where we know the load impedance and wish to derive the input impedance. In practice we are at least as likely to want to be able to derive the load impedance from a known input impedance. If we go back to the skeleton equation (Equation 18) we can transpose it to:

$$\rho_A = \rho_{IN} / k \qquad \text{(Eq 20)}$$

We can take a shortcut to the transposed transmission line equation. Note that in going from Equation 18 to Equation 20, ρ_A and ρ_{IN} have changed places with each other and k has become $1/k$. If we incorporate these substitutions into Equation 19 and multiply both numerator and denominator by *k* we get:

$$Z_A = Z_0\left[\frac{k(Z_{IN} + Z_0) + (Z_{IN} - Z_0)}{k(Z_{IN} + Z_0) - (Z_{IN} - Z_0)}\right] \qquad \text{(Eq 21)}$$

If you care to work it through using the values for our length of RG-213 you will find that it gives the expected answer.

Summary

An alternative transmission line equation has been presented that uses neither exponential nor hyperbolic functions. This has the obvious advantages of making it more usable on pocket calculators and more understandable to anyone whose math education stopped below university level.

The mathematical elegance of the hyperbolic transmission line equation is undeniable and its derivation is based on the same basic principles as the alternative equation presented here. The starting point is the skeleton equation (Equation 18) but with *k* substituted by its exponential function (see Equation 17). The relationship between *k* and its exponential function is based on an amazing equation known as Euler's formula.[3,5] The subsequent manipulation to the hyperbolic equation is also based on relationships derived from Euler's formula.

It is worth noting here that the need to use the hyperbolic function arises from taking account of line loss and if this is ignored the equation becomes:

$$Z_{IN} = Z_0\left[\frac{Z_A + jZ_0\tan\theta}{Z_0 + jZ_A\tan\theta}\right] \qquad \text{(Eq 22)}$$

This simplified equation can be handled on an ordinary scientific calculator and the arithmetic is less daunting. But unless the loss figure is very low, accuracy is seriously compromised.

In this age of computers, ease of calculation need no longer be an issue because anyone who is interested can avail themselves of software which does all the work. But the hyperbolic transmission line equation dates back to the era of log tables and slide rules when any complex algebra calculations must have been very tedious and time-consuming. My guess is that in those days, the Smith Chart would have been the method of choice and the transmission line equation would have been rarely, if ever, used.

That does not, however, explain why the originators of the transmission line equation should have opted for the exponential/hyperbolic solution when there was an easier-to-understand, albeit less mathematically elegant, alternative. The only reference that I have seen to the origin of the transmission line equation is that given in *The ARRL Antenna Book* which attributes it to Ramo, Whinnery and van Duzer.[6,7] I suspect that to them, the exponential/hyperbolic solution would have been no more difficult than the alternative presented here.

Notes
[1] R. Chipman. *Theory and Problems of Transmission Lines*, p 130, Schaum's Outline Series, McGraw-Hill Book Co, 1968.
[2] S. Best, VE9SRB, "Wave Mechanics of Transmission Lines Part 1, Equivalence of Wave Reflection and the Transmission Line Equation," *QEX*, Jan/Feb 2001, pp 3-8.
[3] M. Boas. *Mathematical Methods in the Physical Sciences*. 2nd Edition. John Wiley & Sons, 1996, Chapter 2.
[4] R. Barker, G4JNH, "Improved Remote Antenna Impedance Measurement," *QEX*, Jul/Aug 2004, Appendix 1, p 42.
[5] Dr Math Web site. **mathforum.org/drmath/faq/faq.euler.equation. html**.
[6] R. D. Straw, N6BV, Editor, *The ARRL Antenna Book*, 17th edition, pp 27-29.
[7] S. Ramo, J. Whinnery and T. Van Duzer, *Fields and Waves in Communications Electronics*. Chapter 1, John Wiley and Sons, 1967.

Appendix A: Hyperbolic Tangent Expansion

Ordinary scientific calculators, including those capable of handling complex numbers, will not compute the hyperbolic tangent of a complex number. So unless you have one of the very sophisticated calculators that will, it's necessary to expand tanh into an expression that can be handled. To show why it's desirable to avoid the exercise, here's the math.

For a real number x, Euler's formula can be used to derive the following relation:

$$\tanh x = -j \tan\left(jx \right)$$

$$= \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$= \frac{e^{2x} - 1}{e^{2x} + 1}$$

(Eq A1)

For a complex number $x + jy$:

$$\tanh\left(x + jy \right) = \frac{\tanh x + \tanh jy}{1 + \left(\tanh x \right)\left(\tanh jy \right)}$$

(Eq A2)

The first step to solution uses the identity: $\tanh jy = j \tan y$. Now we have:

$$\tanh\left(x + jy \right) = \frac{\tanh x + j \tan y}{1 + j \left(\tanh x \right)\left(\tan y \right)}$$

$$= \frac{\tanh x + j \tan y}{1 + j \left(\tanh x \right)\left(\tan y \right)} \left[\frac{1 - j \left(\tanh x \right)\left(\tan y \right)}{1 - j \left(\tanh x \right)\left(\tan y \right)} \right]$$

$$= \frac{\tanh x \left(1 + \tan^2 y \right) + j \tan y \left(1 - \tanh^2 x \right)}{1 + \left(\tanh^2 x \right)\left(\tan^2 y \right)}$$

(Eq A3)

The equation can be simplified further using the identities:

$$\tanh^2 x = \frac{\sinh^2 x}{\cosh^2 x} = \frac{\cosh 2x - 1}{\cosh 2x + 1}$$

(Eq A4A)

$$\tan^2 y = \frac{\sin^2 y}{\cos^2 y} = \frac{1 - \cos^2 y}{\cos^2 y}$$

(Eq A4B)

The simplest result is:

$$\tanh\left(x + jy \right) = \frac{\sinh 2x + j \sin 2y}{\cosh 2x + \cos 2y}$$

(Eq A5)

You still have to compute sinh and cosh using:

$$\sinh 2x = \frac{e^{2x} - e^{-2x}}{2} = \frac{e^2}{2}\left(e^x - e^{-x} \right)$$

(Eq A6A)

$$\cosh 2x = \frac{e^{2x} + e^{-2x}}{2} = \frac{e^2}{2}\left(e^x + e^{-x} \right)$$

(Eq A6B)

but at least it can be done on a regular calculator. Isn't it nice to be rid of such grunge?
— *Doug Smith, KF6DX,* QEX *Editor*.

Appendix B — A Worked Solution of the Hyperbolic Transmission Line Equation

This is the hyperbolic transmission line equation:

$$Z_{IN} = Z_0 \left[\frac{\dfrac{Z_A}{Z_0} + \tanh(\gamma L)}{1 + \dfrac{Z_A}{Z_0}\tanh(\gamma L)} \right] \tag{Eq B1}$$

where:
Z_0 = characteristic impedance (Ω)
Z_A = load impedance (Ω)
Z_{IN} = input impedance (Ω)
L = line length in meters.
γ = line propagation factor ($\alpha + j\beta$)
and:
α = line loss in nepers per meter.
β = angular electrical length of line in radians per meter.

The following is the solution of the equation for input impedance based on the parameters listed below:
Characteristic impedance, Z_0 50.0 – j0.3 Ω.
Length   30.0 meters.
Matched loss   0.8 dB.
Velocity factor   0.66
Frequency   14.2 MHz.
Load impedance, Z_A 35.0 – j25.0 Ω.

Ordinary scientific calculators including those capable of handling complex numbers will not give the hyperbolic tangent of a complex number. So unless you have one of the very sophisticated calculators which will, it is necessary to expand tanh into an expression that can be handled. There are several options available; the preferred one being:

$$\tanh(x + y) = \frac{\sinh 2x + j\sin 2y}{\cosh 2x + \cos 2y} \tag{Eq B2}$$

Applying this to Equation B1 we get:

$$\tanh(\gamma L) = \tanh(\alpha L + j\beta L) = \frac{\sinh 2\alpha L + j\sin 2\beta L}{\cosh 2\alpha L + \cos 2\beta L} \tag{Eq B3}$$

We must first establish the values of αL and βL. Note that αL is the line loss in nepers which are the natural logarithm of the voltage multiplier. The relationship of nepers to dB is:

nepers = dB $\times$ 0.11513 $\qquad\qquad$ (Eq B4)

So the loss of the 30 meter length of RG-213 is:

$\alpha L = 0.8 \times 0.11513 = 0.092104$ nepers
βL is the line length in radians and we can determine this as follows:

$$\beta L = \frac{2 \times \pi \times \text{actual length (meters)} \times \text{frequency (MHz)}}{\text{speed of light in free space (Mm/sec)} \times \text{velocity factor}} \tag{Eq B5}$$

So, for the 30 meter length of RG-213 the line length in radians is:

$$\beta L = \frac{2 \times \pi \times 30.0 \times 14.2}{299.8 \times 0.66} = 13.5274 \text{ radians}$$

We can now enter these values into Equation B3:

$$\tanh(\gamma L) = \frac{\sinh(2 \times 0.092104) + j\sin(2 \times 13.5274)}{\cosh(2 \times 0.092104) + \cos(2 \times 13.5274)} \tag{Eq B6}$$

$$\tanh(\gamma L) = \frac{0.18525 + j0.93894}{1.017 - 0.3441}$$

$$\tanh(\gamma L) = \frac{0.18525 + j0.93894}{0.6729}$$

$$\tanh(\gamma L) = 0.2753 + j1.3957$$

In Equation B1, two complex divisions can be avoided by multiplying both the numerator and the denominator inside the brackets by Z_0. Then:

$$Z_{IN} = \frac{Z_0 \left[Z_A + Z_0 \tanh(\gamma L) \right]}{Z_0 + Z_A \tanh(\gamma L)} \qquad \text{(Eq B7)}$$

Inserting the RG-213 values we get:

$$Z_{IN} = \frac{(50.0 - j0.3) \times \left[35.0 - j25.0 + (50.0 - j0.3) \times (0.2753 + j1.3957) \right]}{50.0 - j0.3 + (35.0 - j25.0) \times (0.2753 + j1.3957)}$$

$$Z_{IN} = \frac{(50.0 - j0.3) \times (35.0 - j25.0 + 14.184 + j69.702)}{50.0 - j0.3 + 44.528 + j41.967}$$

$$Z_{IN} = \frac{(50.0 - j0.3) \times (49.184 + j44.702)}{94.528 + j41.667}$$

$$Z_{IN} = \frac{2472.6 + j2220.3}{94.528 + j41.667}$$

$Z_{IN} = 30.6 + j10.0 \ \Omega$.
— *Ron Barker, G4JNH, VK3INH*

Observing Selective Fading in Real Time with *Dream* Software

The author uses digital shortwave broadcast signals to study ionospheric conditions and their effects on selective fading of the signals.

John Stanley, K4ERO

[*For a more technical description of the digital audio broadcasting standards discussed in this article, see C. Demeure and P.-A. Laurent, "International Digital Audio Broadcasting Standards..."* QEX, *Jan/Feb 2003;* **www.arrl. org/tis/info/pdf/x0301049.pdf**. — *Ed.*]

Anyone who has used a radio of any kind is aware of fading. Signals are not of a constant strength, but vary with time, sometimes giving good reception and later, being weak or even unusable. There are quite a number of reasons why signals fade. One type of fading that is of special interest is called "selective fading."

Selectivity means frequency discrimination, and selective fading refers to the case where fading is a function of frequency. Transmission loss on one frequency may be low while on a nearby frequency, it is much greater. This phenomenon can cause strange results. Understanding it can lead to strategies to reduce the effect of this fading on radio communications. Recently available tools can help us to understand the phenomenon.

DRM and *Dream*

Both analog and digital transmissions are affected by selective fading. The advent of digital sound transmissions on the HF band opened up a new way to observe and study selective fading. The free *Dream* software, intended for reception of DRM (Digital Radio Mondiale) broadcasts, includes analysis tools that actually allow us to observe the effects of selective fading in real time, and to plot those

---

524 White Pine Ln
Rising Fawn, GA 30738
jnrstanley@alum.mit.edu

effects for analysis.[1] I have been recently using *Dream* to observe the condition of the path between Sackville, Canada, and my location near Chattanooga, Tennessee.

For the method to work, we need to have a DRM broadcast, since the *Dream* analysis software will only work when synchronized to a DRM transmission. The Canadian Broadcasting Company (CBC) has been broadcasting to the US with DRM programs for some years. Each evening, starting about 4 PM local time, I can hear the programming for several hours.

Decoding DRM signals requires that an intermediate frequency of about 12 kHz

[1]Notes appear on page 22.

be brought out of a receiver. The Ten-Tec RX320D is one receiver that has such an output. On the DRM Receiver Modifications Web site (**www.drmrx.org/receiver_mods. html**) there is a user's forum that lists other receivers suitable for DRM reception, along with the modifications needed. In addition, I have built a very simple direct-conversion receiver that receives the CBC DRM signals well. It was used for recording the graphs included in this article.

The 12 kHz IF signal is connected to the input of a computer sound card. About 1 GHz of processor speed is needed. In the *Dream* opening screen, select the Evaluation Dialog, then the Channel Transfer Function. If you are
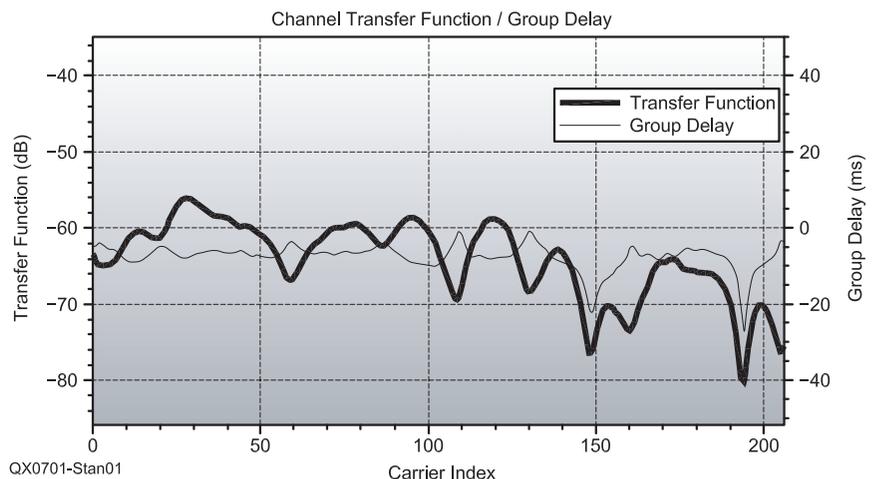
Figure 1 — This screen shot from the *Dream* software shows a signal with moderate selective fading.

locked onto a DRM signal, the screen should look like Figure 1. What we are seeing is the "total path gain" of each frequency within a 10 kHz bandwidth. The DRM signal consists of about 200 carriers spaced about 50 Hz apart, so that the available 10 kHz channel is filled with carriers. The plot shows the relative strength of each carrier as received. Since the carriers are all of identical average amplitude when transmitted, the relative strength of each one at the receiver indicates which frequencies have faded, relative to the rest of the frequencies. Thus we have a dynamic real-time picture of selective fading in action!

Figure 1 shows a typical situation, with moderate selective fading evident. The weakest carrier is about 24 dB weaker than the strongest. (The thick line is the one to read; the thin line shows relative group delay for each channel, which is not of immediate interest.) As you watch the display in real time, you will notice that each sweep, taken less than 0.5 s after the last, shows a quite different pattern, indicating that each channel fades rapidly and the locations of the deep nulls are constantly changing. A waterfall display will allow you to plot the shifting of the deep nulls versus time. Such a display no doubt contains a great deal of information about the state of the ionosphere and merits further observation.

Path Analysis

To verify what we are seeing on the Channel Transfer Display, I did some analysis of the path between Sackville and my location, using both *VOACAP* and also simple geometric analysis. Figure 2. shows the geometry of the path.

As shown in the drawing, the signals from Sackville to my location can arrive by two distinct paths. One, called the 1F mode, involves a single reflection from the ionosphere at the midpoint of the path. The second, the 2F mode, bounces off the ionosphere at two places corresponding to 25% and 75% of the total distance, with a ground reflection at the halfway point. If these two signals arrive at the receiver with nearly equal signal strength, cancellation can occur, causing deep fades.

I downloaded an *ionogram* (**digisonde. haystack.edu/latestFrames.htm**) from the Millstone Hill Web site in Massachusetts, near the path midpoint, to verify the ionosphere's height at the time that the data were taken using *Dream*. See Figure 3. Table 1 was generated by *VOACAP* analysis. The ionogram verified that the MUF was high enough to support both the 1F and 2F modes and that the F layer height was about 220 km. *VOACAP* analysis predicted that both modes were likely present, and that their signal strengths would be nearly equal, which is vital for deep nulls to occur.[2] It also agreed pretty well with the F-layer height figure. *VOACAP* also verified the path length for the two modes, and thus the expected delay for the two paths.

On a path that is long compared with the length of the wave, as the frequency changes, the number of cycles the wave has completed, and thus its phase, will change rather rapidly. If two paths of equal length were involved, this would have no importance; but with unequal path lengths, changing frequency would cause the phase of one signal to shift compared with the other. If the two signals were of equal amplitude, at some frequencies the two signals would be 180° out of phase, and cancellation would occur.

We can predict the frequency difference between adjacent nulls by determining what frequency change would produce a 360° relative phase shift, so as to carry us from

QX0701-Stan02

2228 km Via 1F Mode
Delay = 7.42 µs (74.000 Wavelengths)

2356 km Via 2F Mode
Delay = 7.85 µs (78.500 Wavelengths)

F Layer Height, 220 km

2150 km Along Ground

20°

7°

Sackville, ON

K4ERO
Rising Fawn, GA
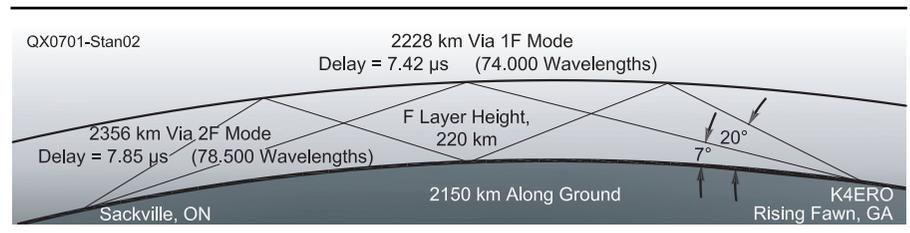
Figure 2 — Signals from Sackville, ON, can arrive at K4ERO in Rising Fawn, GA by one or two ionospheric hops.

Table 1

VOACAP Analysis of Propagation Path

URSI Coeff(Daily) *~METHOD 25* *VOACAP 05.0119W* *PAGE 1*
Nov,06 1994 SSN = 30.
Sackville K4ERO AZIMUTHS N. MI. km

| 45.53 N | 64.19 W | –34.83 N | 85.43 W | 243.87 | 50.01 | 1161.3 - 2150.5 |

XMTR REC705 #01[default\CCIR.012] Az = 243.9 OFFaz = 360.0 50.000 kW
RCVR REC705 #01[default\CCIR.017] Az = 50.0 OFFaz = 0.0

SUMMARY 7 MODES FREQ = 9.8 MHZ UT = 21.0

	1. E	1. E	1. F2	2. F2	2. E	3. F2	3. E	1. F2
TIME DEL.	7.28	7.34	7.47	7.92	7.46	8.91	7.69	7.47
ANGLE	1.60	3.79	7.51	20.39	11.23	33.16	17.96	7.51
VIR. HITE	122.70	165.67	240.32	230.36	131.95	253.85	128.76	240.32
TRAN.LOSS	162.27	169.77	121.75	117.22	450.19	132.42	575.44	121.75
T. GAIN	–3.86	3.91	9.42	14.72	12.25	12.52	14.48	9.42
R. GAIN	–8.24	–0.44	5.22	11.70	8.27	11.83	11.16	5.22
ABSORB	6.76	6.42	5.52	3.11	4.62	2.12	3.42	
FS. LOSS	119.05	119.13	119.28	119.79	119.26	120.80	119.53	
FIELD ST	19.99	4.68	47.04	45.10	–284.45	29.77	–412.58	49.24
SIG. POW	–115.28	–122.78	–74.76	–70.23	–403.20	–85.43	–528.45	–68.82
SNR	47.23	39.72	87.74	92.28	–240.70	77.08	–365.95	93.68
MODE PROB	0.58	0.58	1.00	0.95	0.00	0.63	0.00	1.00
R. PWRG	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1.15
SIG LOW	25.00	25.00	19.18	20.81	25.00	25.00	19.12	20.39
SIG UP	18.22	25.00	7.68	7.76	25.00	10.20	7.68	7.81

one null to the next. From Figure 2, note that the difference between the two path lengths is about 4500 wavelengths at 9.8 MHz: the frequency of the CBC transmission. That would mean that a change of 1 part in 4500 in frequency would produce the next null. That represents about 2 kHz. Thus we would expect that within the 10 kHz span that we are observing, we could expect to see multiple nulls. This is indeed the case in many of the observations, including Figure 1.

Dream has another analysis mode, called the Channel Impulse Response. This shows what would be received if a very short pulse were transmitted, and indicates the time-domain response of the channel. In Figure 4A, we have an example where the 1F and 2F signals are nearly equal. This would produce

What Does an Ionogram Show?

The X axis represents frequencies at which the vertical sounding was done. The Y axis is proportional to the time delay of the return echo(s), but is plotted directly as distance.

This represents the "virtual height" of the reflecting layer. The thin black curve is a plot of electron density versus height. For this thin black curve, the Y axis is again the height above ground and the frequency plot shows the frequency that would be reflected as an O wave by the electron density at that height. Thus the peak of the black curve gives us the maximum frequency at which the F2 layer would reflect a vertically incident wave, fo_{F2}.

Above this frequency there will be a skip zone, although the X wave will provide some signal. Above the frequency where the X wave bends upward there will be a totally dead skip zone. Only ground wave or backscatter signals will be heard.

For those unfamiliar with ionograms, the tutorial at the following Web site is recommended: **www.ngdc.noaa.gov/stp/IONO/Dynasonde/whatis.htm**. An older and much longer treatment may be found at: **www.ips.gov.au/IPSHosted/INAG/uag.htm**.
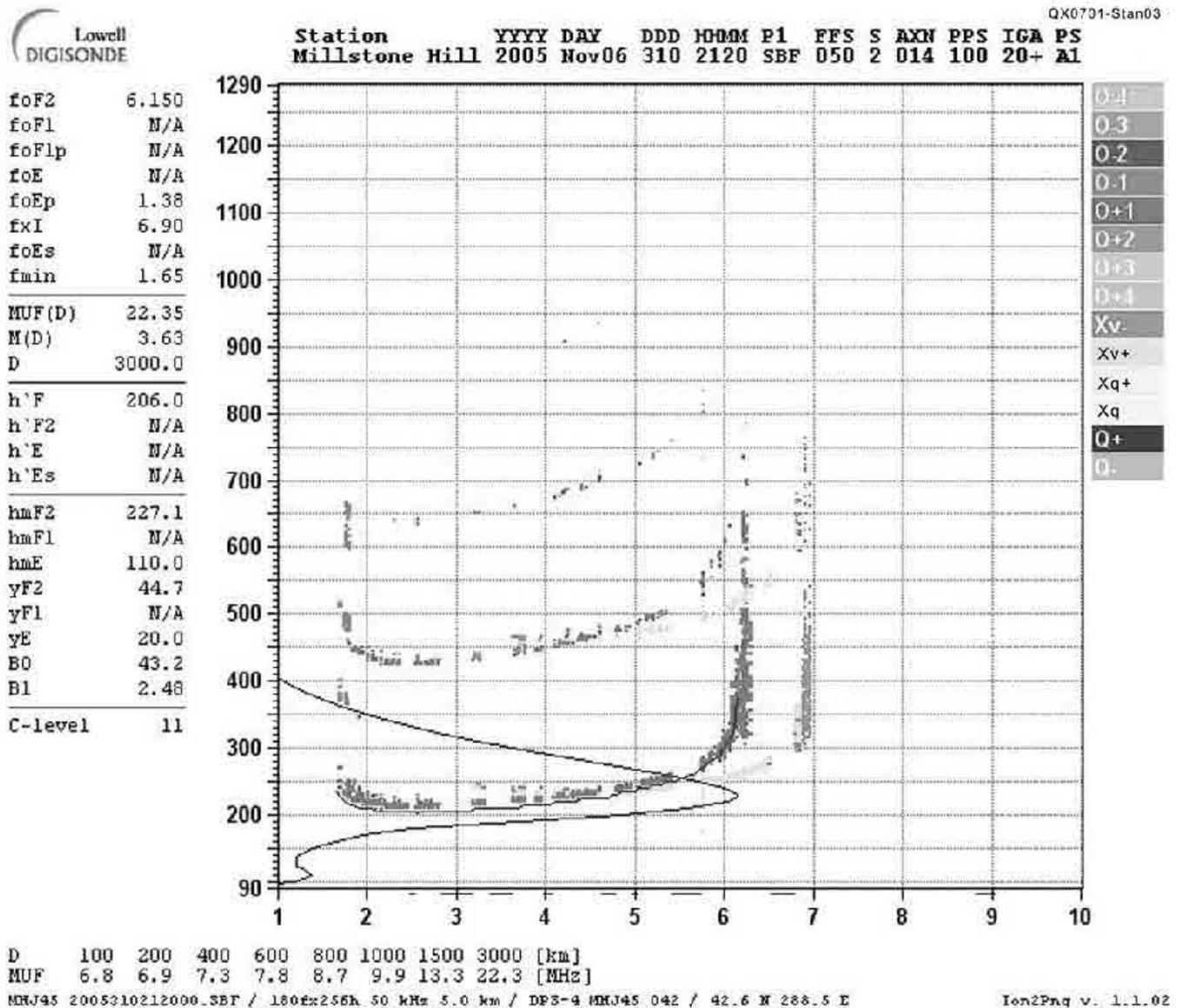
Figure 3 — This ionogram was downloaded from the Millstone Hill Web site on Nov 6, 2005. The graph confirms that the F layer height is close to 220 km, and the two lines of text at the bottom confirm that the MUF for a 1000 km hop is 9.9 MHz. You can find the latest ionogram taken by the Millstone Hill Observatory at **digisonde.haystack.edu/latestFrames.htm**.
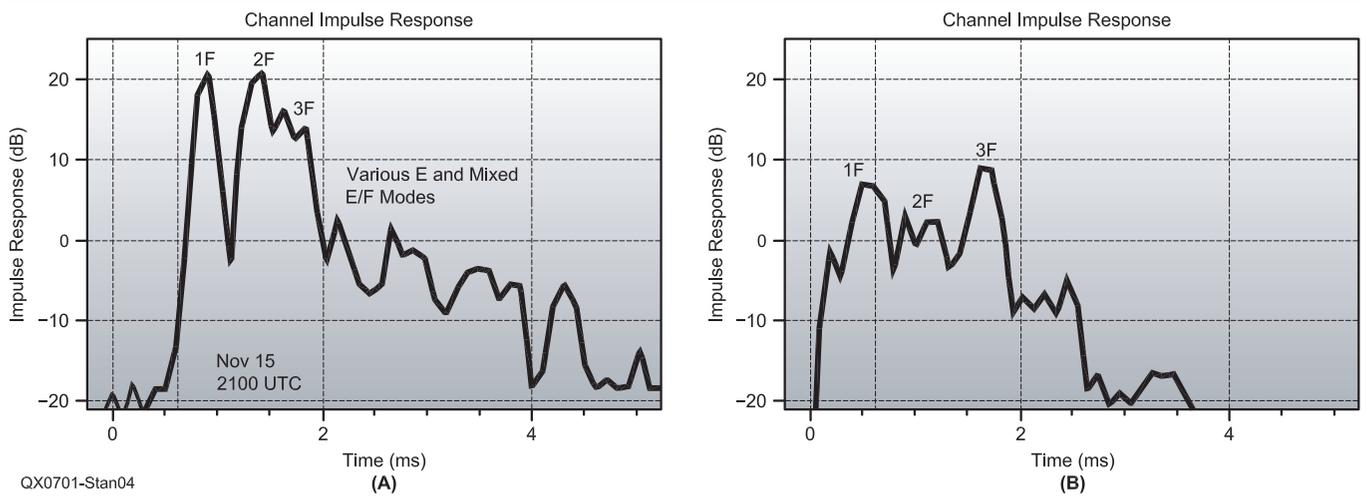
Figure 4 — Part A is the *Dream* software channel impulse response for a case when the 1F and 2F paths are nearly equal in amplitude. This pattern shows a 3F path as well as various mixed E and F propagation modes. Part B shows a channel impulse response graph for a case when the 1F and 3F paths are nearly equal while the 2F path is slightly smaller.

a channel response similar to that in Figure 1. In Figure 4B, we see three pulses, where the second one is smaller than the first and third. The first is probably propagated via the 1F mode; the second, slightly smaller impulse is the 2F mode; and the third, the largest, is the 3F mode. The time between the 1F and 3F modes is more than 1 ms, indicating that fades closer than 1 kHz apart would be possible at that time.

Figure 5 shows a case where spacing between nulls is as small as 1 kHz. This would require a differential delay of about 1 ms, and implies that even the 3F mode sometimes equals the 1F mode, as shown in Figure 4B. Looking back at the VOACAP predictions, we find a 63% prediction that 3F is present. It is expected to be considerably weaker than the 1 and 2-hop F modes, but there can be times when it is comparable to them.

Note that the sum of two nearly equal signals beating together looks like a rectified sine wave when plotted on a logarithmic scale. The nulls will be much more pointed than the peaks. This pattern is often observed on the signals from CBC, thus providing high confidence that we are observing selective fading produced by interference between modes, very likely the 1F, 2F and 3F modes. Figure 5 is a clear example of the rectified sine wave pattern.

Thus, we have a method of observing directly on an HF signal a natural phenomenon that has long been the bane of short-wave radio users. Reflections of any signal from the ionosphere will produce similar results. Reflection from other objects can produce selective fading even in the VHF/UHF range, especially when the receiver is in motion.
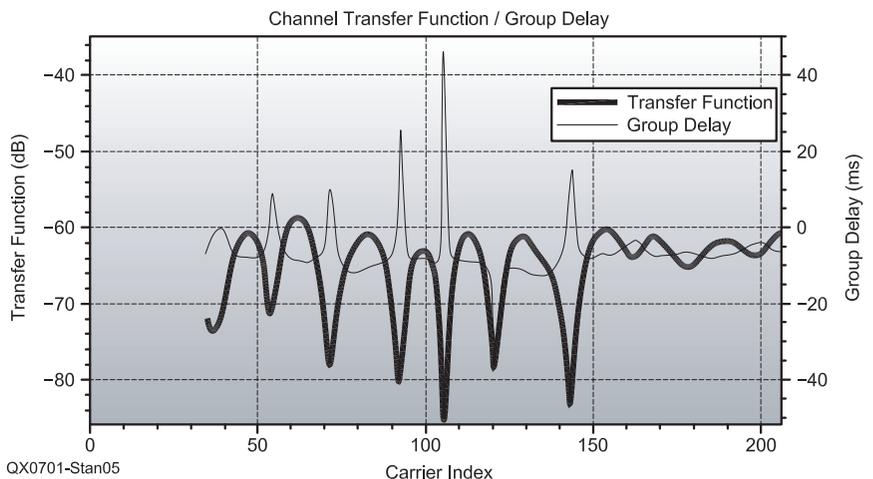
Figure 5 — This graph shows the "rectified sine wave pattern" that results when two nearly equal signals beat against each other.

Selective Fading Effects

Let us consider what deep notches in the spectrum of a received signal could do. For an AM signal, the results often are serious. When the carrier frequency falls into a deep notch, the signal becomes a double-sideband (DSB) signal. If we are receiving it on an SSB receiver, this is of no consequence; but with conventional envelope detection, severe distortion occurs. On the AM broadcast band (540 to 1700 KHz) the effect is usually produced by a beat between the ground wave and the sky wave, and can last for many seconds. These dropouts usually seem to happen just as a long fly ball is headed towards the outfield fence: "Back, back, to the warning track…" During

those seconds, the audio turns to mush and nothing can be understood. On short-wave, the time during which the carrier is in the notch will be much more transient, but will happen more frequently, and thus audio quality is degraded overall, but may not be as obvious as to its cause. The use of SSB improves the situation when only speech is being transmitted. This is because the human voice contains much redundant information content, and a sharp notch somewhere in the range of 300 to 3000 Hz will usually not make the signal unreadable, and may not even be noticed.

CW is an interesting case. In Figure 5, the notches in the transfer function are so narrow that one could visualize a CW signal falling

into one, while its keying sidebands were less affected. This takes out the tone, leaving only keying sidebands. Perhaps the practice of using harder keying (wider sidebands) was found to be useful in the early days of commercial CW traffic systems because the operators were able to copy the clicking sounds even when the tone itself was taken out by a deep selective fade. Thus, to this day, harder keying is recommended for dispersive channels. We associate this harder keying with a crisper sound, but why was it recommended only for fading circuits unless for the above-mentioned reason? This issue is further explained in *The ARRL Handbook for Radio Communications*.[3]

HF digital systems must be designed to cope with selective fading. It is remarkable that during the time that the above channel transfer functions were recorded, the DRM audio was not damaged. The DRM coding schemes, having been designed for HF paths, do remarkably well in coping with such a hostile channel. Hams have long observed that with RTTY using FSK, one of the two frequencies (mark or space) can fade deeply, as observed on the cross pattern oscilloscope display, while copy is unaffected: diversity reception. The choice of wide or narrow shift could be tailored to take into account the expected spacing between

nulls as observed with the *Dream* software, or calculated by geometry.

The CW and RTTY cases mentioned are examples of mitigating the effects of selective fading, perhaps done without a full understanding of the problem, but simply based on operating experience. Another example of a method used to reduce selective fading would be the control of antenna patterns.

This has long been done on the AM band with the use of "anti-fading antennas," which increase the ground wave and suppress high-angle sky waves. Traveling-wave antennas such as rhombics also tend to exhibit less fading than standing-wave antennas such as dipoles.

Note from Figure 2 that the 1F and 2F rays do not arrive at the receiver at the same vertical angle. If one were to install the antenna at a height that would emphasize the 1F signal (7°) but have a notch at the arrival angle of the 2F signal (20°), that would greatly reduce the depth of the nulls. Alternatively, one might use a lower antenna that would improve the 20° gain but reduce the gain at 7°. Hams with stacked beams and the ability to choose which one(s) are fed, or those with adjustable-height towers, have no doubt empirically found that some heights work better than others for a given path. It would be interesting to use such antennas for reception of the 9.8-MHz signals from Sackville and observe changes in the transfer function — especially depth of nulls — as the vertical angle of the array is changed.

Operating near the MUF is another way to reduce selective fading effects. The above

data were taken in early evening. At later hours we would expect the MUF to drop to the point where the 2F mode would no longer propagate. This should be noticeable on the transfer function. Or, we could operate at a time when absorption was greater, so that the 2F mode, with its extra trips through the ionosphere, would be more heavily absorbed.

Notes
[1]A number of sites provide compiled *Dream* software. A Google search on the words "*Dream*, DRM, software, download" should enable one to find the latest version. You can also find more information at **sourceforge. net/projects/drm**.
[2]*VOACAP* software information is available at **www.arrl.org/tis/info/pdf/Voacap.pdf**.
[3]M. Wilson, Ed., *The ARRL Handbook for Radio Communications*, 2007 edition, p 11-7

John Stanley, an ARRL Technical Advisor, was first licensed over 50 years ago as KN4ERO. He has since been active on many bands and from a dozen foreign countries. Now an Extra class licensee, he operates mainly on 60 meter and 2 meter SSB when at home on Lookout Mountain in northwest Georgia. After graduating from MIT with a BSEE in 1962, John did six years of graduate studies in Theology and Foreign Languages. Over a 40 year career, he taught at four universities, designed, installed and maintained broadcast equipment and worked for Texas Instruments and the UCLA Ionospheric Observatory. He and his wife Ruth, WB4LUA, still consult and do training with many broadcast stations, large and small. **QEX**

In Search of New Receiver-Performance Paradigms, Part 2

We've had a look at enemies of receiver performance from within. Now let's explore the enemies from without.

Doug Smith, KF6DX

In Part 1 of this article, we looked at those inherent characteristics of receivers that limit their performance in the absence of external stresses.[1] Here in Part 2, we go outside the receiver to examine what external conditions limit its performance.

The Enemies from Without

We may divide our definition of a receiver's operating environment into two areas: 1) its external electrical environment, including external electromagnetic field strengths, signals at the antenna port and power-supply stability; and 2) its external physical environment, including temperature, vibration, shock and moisture. Historically, only conditions at the antenna port get scrutiny in routine receiver testing; but the others deserve attention.

Environmental Conditions

Electromagnetic Compatibility (EMI/RFI)

Transmitted RF getting into the audio circuits of a receiver at a co-located or full-duplex installation would normally be discovered in the course of existing test programs; but often, 50 and 60-Hz susceptibility would not. The widespread use of aluminum and plastic in the construction of modern rigs often leaves them vulnerable to external fields from power supply transformers. Tests for that are still a part of MIL-STD-461E and other standards. Electromagnetic susceptibility needs more attention in receiver testing.

A simple test: Put your receiver atop

[1]Notes appear on page 30.

225 Main St
Newington, CT 06111-1494
kf6dx@arrl.org

your linear power supply in the vicinity of the transformer and see if you get any FM at the power-line frequency. You might be surprised. The external field may couple to inductors in the frequency-determining circuits — the usual mechanism — to produce an objectionable effect.

Hams living close to high-power broadcast transmitters know about susceptibility problems. RF can get into their telephones, computers, appliances and yes, internal receiver circuits by paths other than the antenna. Even lower-frequency energy from power lines and other sources has been known to cause problems. One company for which I did some work recently fell into the pitfall. Among other pursuits, they sell and service credit-card machines and one of their clients is a national photo studio chain. That client installed a new anti-shoplifting system that evidently generates fairly strong electromagnetic fields at the checkout counter. Some of the credit-card machines wouldn't function at all within about six feet of the counter! It turns out those units had little or no shielding. Their enclosures were almost entirely injection-molded plastic.

Many of the cables being used with those machines were also unshielded — mostly twisted-pair types. Unshielded external cables act as antennas that conduct external electromagnetic energy inside the equipment to which they're connected. Receivers have power leads, control cables and possibly external speaker or headphone cables and those are potential conduits for harmful energy. European CE (*Conformité Européenne*) testing includes an electromagnetic susceptibility test that calls for all standard cables to be connected.[2] CE testing also includes electrostatic discharge (ESD) and power-supply transient tests.

The testing applies to virtually all electronic equipment sold and marketed within the European Union. Early on, many pundits anticipated that the CE standards would be adopted worldwide, but it hasn't happened. Perhaps we could at least adopt some susceptibility tests. The FCC has done nothing in 47 CFR 15 to address the issue, even after making a significant study of telephone-interference complaints, of which they get over 25,000 per year.

Vibration and Shock

Focus again on those inductive elements of frequency-determining circuits in receivers that I mentioned before. Vibration and shock may induce FM, resulting in what are generally called "microphonics." Physical movement of inductors and other components may induce detectable frequency changes. That can be especially troublesome in mobile equipment.

Vibration and shock are also of interest in the reliability of units. The main concern here is what happens in shipping. UPS and other carriers set standards for packaging so that the potential for damage in transit is minimized and so that insurance claims can be upheld. Double-boxing and suitable internal supports are mandated.

Temperature

Tests for frequency stability, but little else, over particular temperature ranges are required parts of certain commercial test standards for receivers.[3] Other performance parameters can vary significantly with temperature. The implication in manufacturer's specifications is that all specified parameters will be met over the specified operating temperature range.

Liquid-crystal diplays (LCDs) stand out among components that are particularly sen-

sitive to temperature. As they get cold, their visual response times lengthen considerably. So-called "super-twist" graphic LCDs require a backplane voltage of about –22 V dc that is used to control contrast. That type of liquid crystal exhibits a positive temperature coefficient in the backplane voltage that maintains constant contrast. Cold-cathode fluorescent backlights require more voltage to start them when they're cold than at room temperature. LCD manufacturers recommend a 30% margin in start voltage for that reason.

Waterproofing

This is chiefly an issue for handheld transceivers and scanners. To me, the term waterproof means submersible. Amateur Radio gear that can operate in the rain would more properly be called splash-proof or drip-proof. I would think it's worth testing when offered by a manufacturer as a feature. Not all emergencies occur in dry weather and waterproofing goes to reliability.

Power Supply

Receiver frequency and other parameters may vary significantly with power supply voltage. It's of concern, for example, whether a 13.8 V dc-powered receiver continues to operate gracefully as the supply voltage sags toward 10 V dc.

A receiver also needs to be checked under overvoltage and reverse-polarity conditions, and for how it reacts to transients on its power supply input. Not all manufacturers specify the acceptable range of input voltages, but ±15% is a reasonable range for testing. For 13.8 V dc equipment, that is 11.7 to 15.8 V dc; for 120 V ac gear, it's 102 to 138 V ac. In many areas of the world, the ac mains rarely

remain within a few percent of its nominal value. AC-powered equipment needs to be checked at both 50 and 60 Hz. Most linear supplies lose significant capacity at 50 Hz and suffer from increased ripple.

Spurious Responses

A receiver spurious response is the reception of a signal on an undesired frequency. The mechanisms by which receiver spurious responses appear depend on architecture. All receivers that translate radio signals in frequency incorporate frequency-mixing schemes, and all practical frequency-mixing schemes are prone to nonlinear effects. Mixers also have image responses. In addition, the use of mixers implies the use of local oscillators (LOs). All those things and more contribute to receiver spurious responses.

Notwithstanding that commutating mixers are inherently nonlinear, nonlinearities in mixers and other receiver components are responsible for intermodulation distortion (IMD): the mixing of two or more signals to produce an undesired on-channel signal. Quite properly, IMD is defined as a spurious response. Not treated much are the harmonics of *single* external signals that appear in a receiver's passband.

RF Harmonic Distortion

Let's say you live close to a 50-kW AM broadcast transmitter whose carrier frequency is 1250 kHz. Even if the broadcast transmitter's third harmonic output were zero, you might have a tough time operating near (1250) ×(3) = 3750 kHz because of cube-law nonlinearities in your receiver's front end. What you might hear there is a distorted version of the broadcast signal that occupies three times

the bandwidth of the fundamental.

A high-pass filter mitigates that problem and most modern rigs have a set of selectable band-pass filter(s) to attenuate AM broadcast and other strong out-of-band signals, but some don't. A measure of such harmonic distortion is but one measure of a receiver's spurious response.

IMD3 Dynamic Range and Intercept Point

Third-order intercept point (IP3) has been used for ages as a figure of merit for receivers. That figure applies to any circuit or system that exhibits cube-law nonlinearity and remains well defined.

The advent of IF digital signal processing (IF-DSP) and digital direct-conversion (DDC) receivers has brought a new set of IMD limitations that cannot be adequately measured using the old techniques. Although modern digital methods generally constitute design improvements, their limitations must be modeled and investigated differently. See "IMD in Digital Receivers" by SM5BSZ in the Nov/Dec issue of *QEX*.[4]

IP3 is not a measurement, but an extrapolation (computation) based on the equation:

$$IP3 = 1.5(IMD3DR) + Noise\ Floor \quad \text{(Eq 1)}$$

where *IMD3DR* is the third-order IMD dynamic range (a ratio in dB), and *Noise Floor* is the noise-floor power (in dBm). In turn, *IMD3DR* is the ratio of the equal power of each of two off-channel tones input to a receiver to the IMD3 product they produce, whose power is equal to the noise-floor power. In a receiver obeying a perfect cube-law nonlinearity, theory says it doesn't matter at what level you measure the IMD, since you

Table 1
A Selection of Receiver IMD3 Test Results from ARRL Product Reviews

All tests are on 20 m with 20-kHz spacing, pre-amp off, except as noted. Noise floor and IMD3DR values are the published values.

Unit	Noise Floor (dBm)	IMD3DR (dB)	IP3 (publ'd) (dBm)	IP3 (calc'd) (dBm)	IP3 Discrepancy (dB)
FT-857	−132	87	+4.1	−1.5	+5.6
IC-756 Pro II	−131	97	+20.2	+14.5	+5.7
TS-2000	−129	94	+19	+12	+7
WJ-1000	−133	97	+30	+12.5	+17.5
Pegasus	−132	77	+7.2	−16.5	+23.7
FT-100	−133	94	+10	+8	+2
FT-1000 Mk5	−127	101	+25.7	+24.5	+1.2
SG-2020	−130	88	+15.5	+2	+13.5
PC-16000A	−127	94	+17.6	+14	+3.6
NRD-545	−135	91	+4.5	+1.5	+3
IC-910H (70 cm)	−142	80	−5.8	−22	+16.2
Yaesu FTDX 9000	−123	101	+35	+28	+7
Orion	−128	95	+23	+15.5	+7.5
Orion II	−127	92	+20	+11	+9

can still relate that reference level to the noise floor by modifying the equation:

$$IP3 = 1.5(IMD3DR) + Reference\ Level$$
$$(Eq\ 2)$$

where *Reference Level* is the observed level of the IMD3 product and *IMD3DR* is the ratio of the equal power of the two off-channel tones to the reference level. We're finding that receivers and other complex systems don't always obey that perfect cube law.[5] Why not?

Table 1 shows a smattering of figures reported over the years. The departure from cube-law is clear. It remains to show whether the measurements have errors, receivers really don't follow a cube law, or both. Note that the discrepancies are all of the same sign.

A single stage in a receiver might well be cube-law, but several stages in series might not be. IMD3 products from each stage may add or subtract depending on their phases. Furthermore, components like crystal filters and analog-to-digital converters (ADCs) have been shown to produce third-order ($2f_1 \pm f_2$) distortions that don't obey a cube law. Those poorly understood effects are the subjects of ongoing study. Finally, automatic gain control (AGC) circuits may come into play.

The details of how an engineer measures IMD3 are well-documented in the *ARRL Product Review Test Procedures Manual* and in *The ARRL Handbook*.[6,7] See Figure 1. Good isolation between the interfering signal sources is essential for accurate measurement. No matter how good the test setup, though, it's possible to design a receiver that fools current procedures. Here's one way.

Arrange for a so-called front-end AGC that reduces the RF gain ahead of the final selectivity element (filter) in the receiver.[8] See Figure 2. Allow the AGC to be activated by signals that are within the *initial* selectivity of the receiver, but outside the *final* selectivity. When strong off-channel signals are present, RF gain is reduced — by an attenuator, for example — as they are for strong on-channel signals, and distortion is mitigated. But receiver noise figure is thereby raised. So, if you measured the noise floor using currently accepted single-signal methods and used it in Equation 1, you'd get an artificially high IMD3DR and IP3. DSP can compensate for the gain variations in your on-channel signal, but your signal-to-noise ratio (SNR) still degrades and IMD3DR and IP3 aren't as high as you think they are. For that reason, noise-floor power (or noise figure) must be measured under the same conditions as IMD3 is measured.

When tested units don't obey a cube law, a cube-law extrapolation like IP3 is *meaningless*. What's needed is a measurement of IMD3 — and possibly a statement of *IP3 equivalent* — of all units under identical, realistic conditions and at more than one reference level. Table 1 doesn't reflect that. Section 5.8.3.12 of Note 6 clearly indicates that my Equation 1 *must* be used for the calculation, but it's obviously never been used properly; Equation 2 is being used instead, which yields results that don't comply with Equation 1. Reporting results without declaring an uncertainty makes comparison of independent tests impossible.

Useful reference levels might be noise floor, S-5 and S-9; but the S meter shouldn't be used to set the reference because that also makes valid comparisons impossible. A fixed input power should be established instead. If S-9 were defined as 50 µV RMS into 50 Ω, then S-9 would be –73 dBm; S-5 would be –97 dBm at 6 dB per S unit.

Those who use receivers to grab signals near the noise floor have one set of criteria; those who operate near S-9 have another. If you elect to measure IMD3 with respect to S-9, you cannot necessarily report IMD3DR
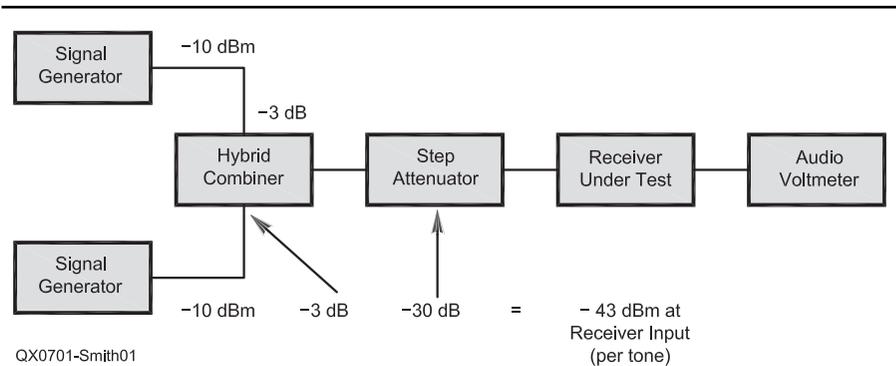
Figure 1 — IMD3 test setup. The audio voltmeter should be replaced with a spectrum analyzer to eliminate "noise-limited" measurements.
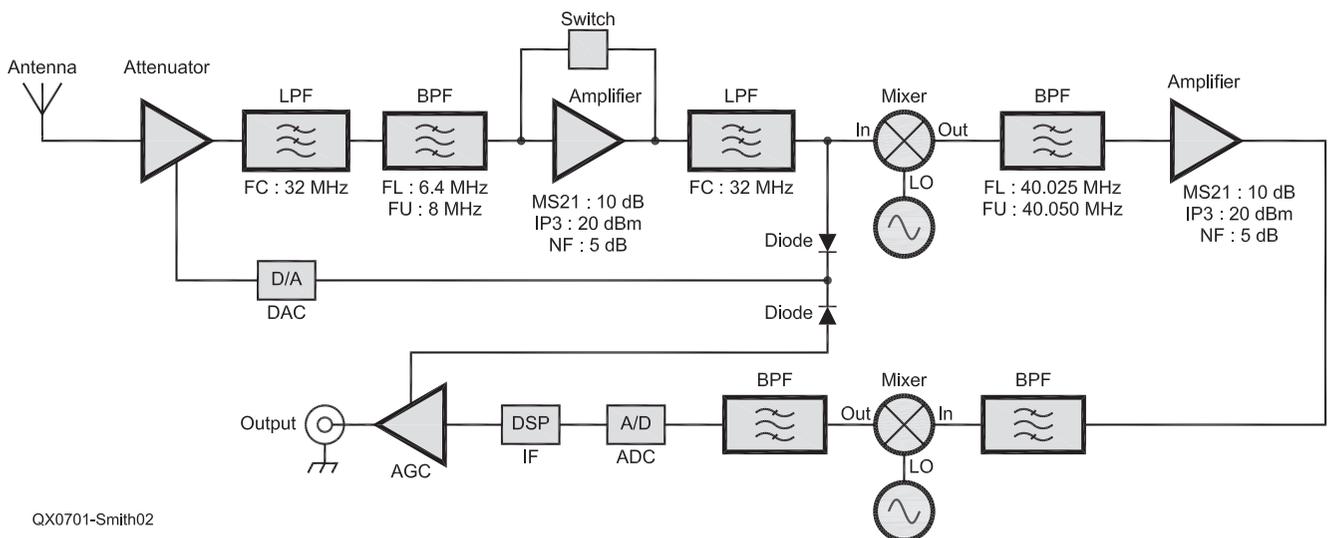
Figure 2 — Block diagram of a receiver having a front-end AGC (from Note 8).

with respect to the noise floor. In other words: If you don't use or measure that part of dynamic range under S-9, you *can* report S-9 IMD3DR equivalent — call it what it is. It's a much smaller number than IMD3DR measured with respect to the noise floor.

If IMD3 performance obeys different laws at two reference levels, how are we to express a single measurement that allows comparison among rigs? When the point at which we measure IMD3 is a moving target, corrections must be made to previously reported data to permit fair comparison. New measurements of old rigs are not always practical and a new definition and model are required. Multiple measurements at multiple reference levels are the only way to characterize a unit's actual performance curve.

When measured IMD3DR is greater with pre-amp on than with it off, it's time for a sanity check. Even if a pre-amp were ideal (IMD-free with a noise figure of 0 dB), IMD3DR would not increase. It would lower the noise floor by the amount of its gain, extending the low end of IMD3DR. It would lower the high end of IMD3DR equally, however, so IMD3DR would remain constant, *but a pre-amp can't increase it*. Any reports of such a circumstance are to be viewed with great disdain unless they're within a declared uncertainty margin. Much more often, low-noise-figure pre-amps have lower IMD3DRs than do receivers they precede, so we should expect IMD3DR to fall with a pre-amp in the circuit.

IMD3DR measurements are sometimes declared to be "noise-limited." What that reveals is that the technician, the instrumentation, or both can't measure what's wanted. See the "Blocking and Phase-Noise Dynamic Range" section below. I'll have more about instrumentation in Part 3.

Half-IF, Image and IF Rejection

A traditional measure of RF second-harmonic distortion, at least in VHF-and-above receivers, is called half-IF rejection. A spurious response lies at one of the frequencies $f_c \pm 0.5 f_{IF}$, where f_c is the tuned center frequency of the receiver. The second harmonic of the LO mixes with the second harmonic of the input signal in the relationship $2 f_{LO} \pm 2 (f_c \pm 0.5 f_{IF}) = f_{IF}$.

Image rejection is a measure of a receiver's response at the image frequency of its first mixer. As an example, let's say you're using a 10.7-MHz IF on a 2-meter receiver. When your receiver is tuned to 146 MHz, the image lies at one of the frequencies given by $f_c \pm 2 f_{IF}$, or 146 ± 21.4 MHz (124.6 or 167.4 MHz). Since those image frequencies are so close to the desired band, elliptical preselectors or notch filters are sometimes used to get good image rejection.

IF rejection is a measure of rejection when the antenna input signal is f_{IF}. The mechanism

here is that the f_{IF} input signal couples directly past any front-end filters and the first mixer, straight into the IF strip.

LO-Related Spurious Responses

Spurs on an LO's output can cause receiver spurious responses because the spur may mix with an off-channel signal to produce an unwanted signal in the passband. For example, let's say you have a spur on your LO caused by a switching power supply in your receiver that's 100 kHz away from f_{LO} and 60 dB below the main LO power. You may get a spurious receiver response 100 kHz away from the tuned frequency that's suppressed only 60 dB — unless your RF preselector is narrow enough to suppress it further.

Such LO spurs may be caused by leakage of internal signals onto the frequency-control elements of the oscillator, or by signals in the frequency-control circuitry itself. In a phase-locked loop (PLL), for example, some energy at the loop's reference frequency is inevitable to keep the loop locked. High reference frequencies in PLLs are desirable because they shorten lock times and because reference spurs are easy to filter. They also imply large frequency steps in simple PLLs and often, 5 kHz or less is the desired step size and also the reference frequency. Suppression of reference spurs needs to be equal to or greater than the desired level of receiver

spurious response rejection.

Harmonics in the LO source are generally not of concern in generating spurious responses because the mixer generates plenty of those harmonics anyway, but they may generate "birdies" as described in Part 1. An exception arises in the case of some DDS circuits. I repeat some of the discussion from Part 1 here for clarity.

Direct digital synthesis (DDS) is a brute-force way of generating a sine-wave LO that uses a numerically controlled oscillator (NCO) coupled to a digital-to-analog converter (DAC). See Figure 3. Its main drawback to date has been spectral impurity. Both discrete and broadband AM and PM impurities may be characterized and analyzed.

Some of the spurs at the output of a DDS are harmonics of the output signal, and the fundamental and harmonics of the clock. Harmonics of the output signal are chiefly caused by nonlinearities in the DAC. Harmonics that fall at more than half the clock frequency may cause unexpected problems because they can produce aliases that fold back to frequencies at less than half the clock frequency.

The DDS-driven PLL has been a popular approach to frequency synthesis at least since the early 1990s. See Figure 4. PM spurs in such a system may be limited to acceptable values of <<–80 dBc in careful designs, because spurs outside the PLL loop bandwidth

QX0701-Smith03
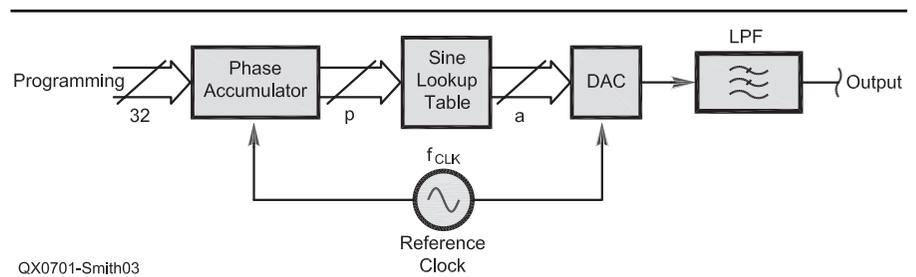
Figure 3 — Block diagram of a DDS.

QX0701-Smith04
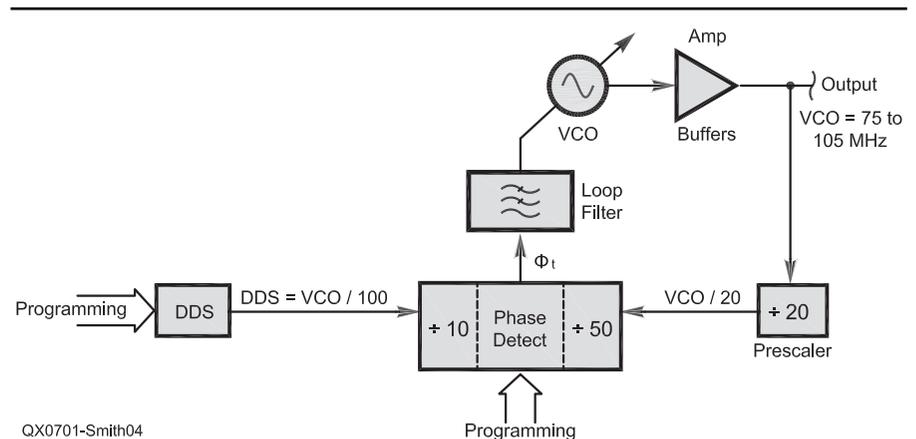
Figure 4 — Block diagram of a DDS-driven PLL hybrid.

are automatically suppressed. Spurs inside the loop bandwidth, however, are *amplified* by the ratio of LO frequency to reference frequency and suppressions of 70 to 90 dB are common.[9]

The point is that with all the possible sources of spurious responses, we should spend some time predicting and searching for them in any receiver evaluation. To ignore them is to present an incomplete picture of performance.

Phase-Noise Dynamic Range and Blocking

Blocking has been treated extensively in another *QEX* article, but it's worth a brief mention here.[10] We have to admit that most modern receivers don't have stages that suffer from saturation, which is the traditional source of blocking, before other effects take over. Instead, LO phase noise and the reciprocal mixing phenomenon determine the limits of selectivity at reasonable frequency spacings. We must admit, too, that while insisting on a very low-phase-noise source for phase noise dynamic range testing is prudent, in actual usage a receiver may suffer as much or more from the phase noise on an interfering transmitter than from its own receiver's phase noise.

Incidentally, we might state a parallel caveat for receiver IMD. On-air interference might be determined by the IMD of adjacent transmitters instead of by one's receiver.

Phase-noise dynamic range in a receiver may be measured by existing blocking dynamic range procedures. The important thing is that we change the name of the test to avoid ambiguity. Although in a transceiver, a receiver's phase noise might be expected to be the same as that of the transmitter, the transmitter might add noise that isn't present in the receiver.

Automatic Gain Control (AGC)

AGC attack time is especially important to receiver performance. An on-air signal's CW envelope might be perfect, but a poor AGC can destroy it. Receiver transient responses to large and rapid excursions in signal amplitude need investigation.

Conventional audio and IF-derived AGCs are control systems that face a distinct disadvantage: They cannot act on changes in signal amplitude until after they detect the changes. As in any servo system, proper loop gain, damping and frequency response are critical to optimal performance. DSP-based receivers face unique challenges when it comes to AGC but they have outstanding advantages, too. It's useful to see how both analog and digital AGC systems behave in such receivers because an understanding of them leads directly to possible additional tests, or the need for changes to existing tests, or both.

Today's radio receivers need to handle a tremendous range of signal amplitudes, especially on HF. From a noise floor of about 0.1 μV to a maximum usable signal of around 1 V, the range is 140 dB! Users would like receiver output signals to remain at constant amplitude over most of that range. AGC does the job by setting receiver gain to be inversely proportional to input level.

DSP and its interaction with analog electronics are emphasized in the following. Part of it is a refinement of the digital AGC I've described previously.[11, 12]

IF-DSP and Digital AGC

IF-DSP receivers produced at the time of this writing generally employ both analog and digital AGC systems. That is because the DSP section alone cannot achieve the 140 dB of dynamic range required. Analog AGC kicks in at some high input level to prevent overload of the ADC, thereby extending dynamic range.

At input levels below actuation of analog AGC, digital AGC is solely responsible for leveling output signals. DSP peak-detects IF signals falling within the desired passband and adjusts a digital gain-control factor to maintain constant peak output. Figure 5 is a simplified block diagram of this system, which is identical to that of a traditional analog AGC.

In digital AGC systems, it is relatively easy to provide a variable threshold or "knee." Input signals below the threshold do not actuate the digital AGC and are not compressed. At thresholds well above the receiver noise floor, a receiver therefore gets quiet when only puny input signals are present in the passband. At thresholds near the noise floor, all signals are boosted to meet the output-level criterion. The net effect of a variable threshold is very much like that of an IF gain control.

A good default setting for threshold is about 3 μV. That means you have about 30 dB of linear range between the noise floor and the point at which AGC starts operating. With a variable threshold, you can set the AGC threshold manually. A low threshold

(0.35 μV) means all signals are boosted to a constant peak output level; a high threshold (350 μV) means signals must reach about 17 dB over S-9 before compression occurs.

AGC decay rates describe how quickly IF gain increases in the absence of signals over the threshold. During decay, IF gain increases geometrically with time — that is, by a certain number of dB per second. A slow setting might run about 5 dB/s, while a fast setting is many hundreds of dB/s. An off setting might make decay time very short. Fast and off would be such that the AGC may actually destroy the envelopes of signals in the passband. The net result is clipping, which produces distortion — but you wanted it fast, right?

It's also fairly easy to implement a peak-hold or "hang" function that retains the most-recent peak for an adjustable period of time. The S meter should reflect the behavior of the AGC system in all ways. Attack time is generally fixed, on the order of milliseconds, so as not to respond to narrow noise pulses.

When noise reduction is engaged, it is desirable to artificially reduce the AGC threshold; otherwise, things get very quiet indeed! Because of that, you may notice that your audio level increases as you turn on noise reduction; but signal-to-noise ratio improves and that is the criterion, after all.

Analog and Digital AGCs Together

An IF-DSP or DDC receiver uses digital filtering for its final selectivity. That means its DSP samples more IF bandwidth than what's desired at the receiver output. Very large input signals may actuate its analog AGC, reducing the gain between antenna and DSP. For in-band signals, that's no problem; but if the large signals are outside the final passband, analog gain is also reduced for in-band signals. The receiver's output amplitude will bop up and down as the analog AGC is pumped by the interference.

The general solution is to employ digital gain compensation. To do it, the DSP must have information about the amount of analog
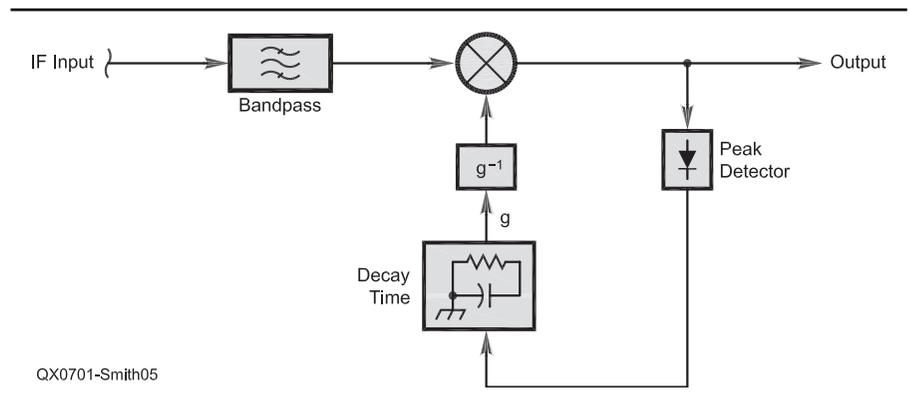
Figure 5 — Block diagram of a traditional AGC system.

gain reduction and the ratio of in-band signal amplitude to interference amplitude.

Digital Gain Compensation

For traditional analog AGC systems not under the control of the DSP, analog gain-reduction information may be obtained by digitizing the AGC voltage. See Figure 6. The voltage value may be used to look up a gain-reduction factor from a table stored in non-volatile memory. Such a table may be built using measurements of the actual hardware. Minor unit-to-unit variations are readily handled by placing the digital gain-compensation point inside the main digital AGC loop, as described below.

An alternate approach involves generating the analog AGC voltage in the DSP itself. See Figure 7. A DAC develops a voltage for application to analog gain-controlled stages. The chief drawback to the scheme is a significant delay between peak detection and gain change, since signals must propagate all the way through the DSP section before being detected. That can be compensated with a delay in the analog IF strip; but typically, the required delays of several milliseconds are impractical. A much better solution is described below under *"Preventing AGC Overshoot."*

In any case, call the analog gain-reduction factor g, where $0 < g < 1$. For example, were $g = \frac{1}{2}$, analog gain reduction would be $-20 \log(\frac{1}{2})$ or about 6 dB. Now it remains for the DSP to compute how much of that gain reduction was caused by in-band signals and how much by interference. If all of it were caused by in-band signals, no gain compensation would be necessary and we would use a digital gain-boost factor $f = 1$. If all of it were caused by interference, in-band signals would have to be boosted by a factor $f = g^{-1} = 2$. For cases in between those two extremes, the procedure is a little tricky because f cannot be described by a single equation.

A Case Study in Gain

To get information about the ratio of in-band signals to interference, the DSP peak-detects both the broadband IF (everything that is digitized) and the receiver output. See Figure 8. Call the peak interference level m and the peak in-band signal n. The peak of the broadband IF is therefore the sum of the interference and in-band signals, or $m + n$. The DSP calculates the ratio: $k = (m + n) / n = m / n + 1$. The next step is to determine whether n by itself was large enough to actuate analog AGC. The DSP does that by comparing k with g^{-1}. The algorithm accounts for three cases in the comparison.

Case 1: If $k < g^{-1}$, then n by itself is large enough to actuate analog AGC and the gain-boost factor used is $f = k$. The ratio of signals solely determines the boost factor.

Case 2: If $k > g^{-1}$, then n by itself is not large enough to actuate analog AGC and the gain-boost factor is $f = g^{-1}$. Analog gain reduction solely determines the boost factor.

Case 3: When $k = g^{-1}$, it obviously does not matter which is used as the gain-boost factor since they are equal. Remember that when analog AGC is inactive, no gain boost need be applied.

Note that g depends only on the characteristics of the analog gain-controlled stage or stages; k depends on the ratio of in-band and interfering signals, irrespective of the analog section. The two possible gain-boost variables therefore produce different functions and curves. The curves are guaranteed to meet where $k = g^{-1}$.

Gain Boost Belongs Inside the AGC Loop

The decay time of the broadband $m + n$ peak detector must match that of analog AGC as closely as possible. The decay time of the in-band n peak detector may be altered at will to get the desired response. Placing the digital gain boost inside the AGC loop assures that a constant peak output level will be maintained even in the face of minor variations in analog gain control. See Figure 9.

Inside the loop, we apply digital gain boost to signals before they are peak-detected, and adjust our computations accordingly. Therefore, the main digital AGC loop prevents

them from exceeding the set output level when interference — and k or g^{-1} — rapidly increase. In addition, IF gain may be manually reduced by artificially increasing the analog AGC voltage without deleterious effects.

Finally, gain-boost factor f may be directly used to compensate a signal-strength meter by the appropriate amount. Just as the receiver output level remains constant in the presence of interference, so does the S meter. When IF gain is manually reduced, the S meter goes down — not up, as in so many rigs. That's also the key to allowing the S meter to reflect the action of notch filters, noise reduction and so forth.

Preventing AGC Overshoot

A DSP normally stores signals to be processed in a series of buffers. Signals from recent to old are therefore available. That presents a neat way of avoiding late adjustment of AGC, or *overshoot*.

When a big signal comes along, receiver input amplitude may rise rapidly from the noise floor to a value of some 100 dB greater, or more. A CW signal with a fast rise time may necessitate a gain-change rate of thousands of dB per second! Digital AGC copes with that by detecting recent signals and applying the gain change to older signals before they are output. In that way, the DSP "sees" the big signal coming before it can destroy a constant

QX0701-Smith06

Figure 6 — Block diagram of a digital system that digitizes a conventional AGC voltage.

QX0701-Smith07

Figure 7 — Block diagram of an alternate system that digitally generates an analog AGC voltage.

QX0701-Smith08

Figure 8 — Block diagram of detectors for digital gain compensation.

output level. The technique need not introduce additional delays in baseband output because the delays are already there and it's only the gain change that moves backward in time.

Signal Conversion

An outstanding goal for digital receivers is to digitize as much bandwidth as possible with as much dynamic range as possible. That goal generally dictates moving the digitization point as close to the antenna as it can be. The ultimate realization, which we're now seeing in practical designs, digitizes amplified and filtered signals directly from the antenna: digital direct conversion (DDC). The present limitations of ADCs, however, still demand the use of frequency conversion and intermediate frequencies (IF-DSP) to obtain maximum dynamic range.

Whether IF-DSP is involved or not, ADCs sometimes determine dynamic ranges of receivers these days. Until very recently, that situation has been something for designers to avoid. IMD3, blocking and other effects, however, don't necessarily dominate the scene as they used to do. Instead, spurious responses come to the fore, along with reciprocal-mixing effects. Spurious-free dynamic range (SFDR) and phase-noise dynamic range (PNDR) are the new criteria.

Nonlinearities in a software radio must be measured, but they are not necessarily cube or square-law: They must be modeled statistically and measured accordingly. New issues arise about comparisons with older designs and methods, but certain limitations that have gone largely undocumented in the Amateur Radio literature remain the same. It will be hard to get rid of the idea of an IP3 equivalent, but ADC makers have never used it.

ADC Distortions

In general, ADCs exhibit two kinds of nonlinearities: differential nonlinearity (DNL) and integral nonlinearity (INL). DNL is a measure of output nonuniformity from one input step to the next. It's expressed in bits as the maximum error in the output between adjacent input steps, over the entire range of possible inputs for the device. Since we're discussing the smallest input steps the converter can resolve, noisy low-order IMD products produced by the effect tend to limit dynamic range.

Manufacturers often specify SFDR for their devices — thank you! Usually, that's given under single-tone conditions; two-tone measurements aren't as common. Harmonic relationships between input signal frequency and the sampling frequency usually determine where discrete spurious responses lie. That can be difficult to illustrate mathematically; but the point is that those responses must be sought and measured by all means.

ADCs are considered *monotonic* if a steady increase in input signal always results

in an increase in output. Backward steps may cause unexpected problems in systems working close to resolution limits. Manufacturers offer different grades of converters that are specified to some number of bits: ±0.5 bits to maintain monotonicity.

INL is a measure of a device's large-signal-handling capability. To test it, we first inject a signal of amplitude *A* and take the output. Then, we inject a signal of amplitude 100 *A* and compare the result with 100 times what we got in the first measurement. We expect the output to increase in exact proportion to the increase at the input. INL is a measure of the output error between any two input levels. Input versus output may be plotted, and maximum deviation from a straight line is then easy to see. That effect obviously produces harmonic distortion and IMD that may be quite undesirable. Typical values for INL center on ±1 bit.

ADCs do not usually exhibit cube-law or even square-law nonlinearities in the same way as do regular analog circuits over their ranges. They do, however, show catastrophic distortion when their input ranges are exceeded: overload. Overload is to be avoided at all costs in a digital receiver and some form of analog AGC is usually employed to avoid it.

As sampling frequencies increase to meet the need of moving the digitization point closer to the antenna, one effect comes increasingly into play: *aperture jitter*. Aperture jitter is noise introduced into ADC results by variations in the exact times of sampling. It's the same as reciprocal mixing because an ADC's clock is, in effect, a local oscillator and it may cause undesired phase modulation of the sampled signal. Assuming that the clock's phase noise is uncorrelated with the input signal, aperture-jitter noise will be uniformly distributed across the sampling bandwidth. We may express the mean aperture jitter in units of time — in seconds. A useful exercise

is to find the noise power caused by the jitter in terms of time. In this case, it's obvious that the signal's frequency will come into the equation because the modulation index will change. In addition, it matters how large the mean aperture jitter, σ_A, is in relation to the sampling period, f_s^{-1}.

I've shown (see Note 11) that the normalized single-sideband noise density (in Hz^{-1}) caused by aperture jitter σ_A (in seconds) is:

$$ND = \frac{4\pi^2 f_c^2 \sigma_A^2}{f_s} \qquad \text{(Eq 3)}$$

Equation 3 assumes a sine-wave input signal. *ND* is the ratio of aperture-jitter power to carrier power, normalized to a 1-Hz bandwidth, where f_c is the carrier frequency and f_s is the sampling frequency. Taking 10 log (*ND*) yields a result in dBc/Hz.

In general, sampling frequency must increase in direct proportion to carrier frequency. Aperture jitter tends to increase in proportion to the carrier frequency, too, and the noise density it produces therefore increases in proportion to the cube of the carrier frequency. That makes it increasingly difficult to maintain dynamic range. When harmonic sampling is employed, the proportions may be different. Much depends on receiver architecture.

Summary

When I did a little research for this part of my article, I didn't discover much about many of the above-mentioned facets of receiver performance in the amateur literature. Perhaps that's because some of them typically aren't as bothersome in amateur work as they are in commercial situations. One reason for that difference is that hams are free to move about anywhere within assigned sub-bands, while commercial operations are channelized. Another reason is that hams are more toler-

QX0701-Smith09

Figure 9 — Block diagram of the hybrid system with gain compensation inside the AGC loop.

ant of obstacles to reliable communications because no money is on the line. When a radio operator aboard ship needs to order more fuel for his vessel, or receive the latest weather information along his route, a lot more is involved than just personal gratification. Still, how a receiver deals with the enemies from without is an indication of equipment quality and is important in buyers' and designers' decisions.

Now you might say that doing all the above tests during product-review testing of every unit is impractical, and you'd be right. What are some appropriate grounds for including tests and possibly excluding others?

Amount of available printed space or some *ad hoc* judgment of the importance of the equipment or the relevant knowledge level of a typical buyer is not enough. A judgment of which tests to apply to a receiver has to do primarily with a balance between intended use, architecture and manufacturer claims. A monoband single-conversion kit might not need extended testing but a radio like the Ten-Tec Orion certainly does. The Elecraft K2 QRP kit does, too, not because it will be used in the same way as an Orion, but because its makers have striven for such a high level of performance in a kit. Hams need to know how well they succeeded and whether they did their testing in the right way.

Certain units are regularly selected for extended testing and that's where all of the additional tests I've suggested should be performed — by reviewers and manufacturers alike. Manufacturers have the opportunity to inform potential buyers about what they do in the way of engineering and production testing, thereby winning additional sales.

I'll have more about specific instrumentation and procedures in Part 3.

Acknowledgments

Thanks again to *QEX* Contributing Editor L.B. Cebik, W4RNL, for editing this manuscript. Thanks to Leif Åsbrink, SM5BSZ, for reviewing it.

Notes

[1]D. Smith, "In Search of New Receiver Performance Paradigms, Part 1," *QEX*, Nov/Dec 2006, pp 23-30.

[2]89/336/EEC, *European Union Council Directive*, 1989; also see IEC1000-4, *Electromagnetic Compatibility*, International Electrotechnical Commission.

[3]47 CFR 2.1055.

[4]L. Åsbrink, SM5BSZ, "IMD in Digital Receivers," Nov/Dec 2006 *QEX*, pp 18-22.

[5]E. Hare, W1RFI, "What is the 'Real' Intercept Point?" in "Improved Dynamic Range Testing," *QEX*, Jul/Aug 2002, pp 50-51.

[6]**www.arrl.org/members-only/prodrev/test-proc.pdf**.

[7]D. Millar, K6JEY, "Test Procedures," Chapter 25, *The ARRL Handbook*, 2007, ISBN: 0-87259-976-0.

[8]For an example, see the receiver design in U. Rohde, KA2WEU, "Theory of Intermodulation and Reciprocal Mixing: Practice, Definitions and Measurements in Devices and Systems, Part 2," *QEX*, Jan/Feb 2003, pp 21-31.

[9]F. Cercas, M. Tomlinson and A. Albuquerque, "Designing with Digital Frequency Synthesizers," *Proceedings of RF Expo East,* 1990.

[10]L. Åsbrink, SM5BSZ, "Blocking Dynamic Range in Receivers," *QEX*, Mar/Apr 2006, pp 35-39.

[11]Chapter 16, *ARRL Handbook*.

[12]D. Smith, KF6DX, *Digital Signal Processing Technology: Essentials of the Communications Revolution*, ARRL, 2001-2003, ISBN 0-87259-819-5.

QEX

The Ultimate Sidetone

Make a vintage radio even better with a modern, well-engineered sidetone.

Mal Crawford, K1MC

Older vacuum-tube and solid-state receivers are available at low cost and can be easily modified for CW operation by adding a sidetone oscillator. Adding a sidetone to a receiver allows the operator to monitor his sending as an aid to maintaining consistent character and word spacing. In many older SSB separate receiver and transmitter pairs, the sidetone oscillator was located in the transmitter so that it could be used as the CW tone oscillator. The sidetone oscillator design described here was incorporated into both the vacuum-tube Heath SB-301 receiver and the later solid-state SB-303 receiver. The Heath SB-303 dates back to 1970, but is still popular because of its good performance, low cost in the used equipment market, and low power consumption.

The Ultimate Sidetone can be used in a wide range of both vacuum-tube and solid-state rigs by changing the values of the gain-setting resistors in the summing amplifier. The summing amplifier component values for both the SB-301 and the SB-303 are provided as reference points for scaling values for other types of receivers. The design uses standard transistors and integrated circuit components that have been featured in the *Hands on Radio* series of experiments in *QST* written by Ward Silver, NØAX.

Previous sidetone oscillators that I built had clicks and pops in the audio when keyed and had rough-sounding rectangular, trapezoidal, or sawtooth shaped waveforms. The design goals for the Ultimate Sidetone oscillator were to have shaped on and off amplitude transitions and a clear, low-distortion sinusoidal tone with low harmonic content. The Exar XR2206 precision waveform generator integrated circuit was selected because its features met those two design goals.

The IC has a sinusoidal signal output in addition to triangle and square wave outputs and an amplitude-modulation port for envelope shaping. The XR2206 also has an output amplitude control and two operating frequencies that can be selected by a logic input. A low-noise operational amplifier was used to sum the sidetone audio without changing the volume-control characteristics. Design goals for the logic portion of the circuitry were to be able to operate with input keying signals from +5 V dc TTL, CMOS, and open-collector logic devices or hand keys, and the ability to disable the oscillator. The 2N7000 MOSFET switching transistors and the LM301A operational amplifier and LM311 voltage comparator integrated circuits in the sidetone oscillator are standard devices that are widely available. A schematic of the Ultimate Sidetone is shown in Figure 1.

The Ultimate Sidetones for the SB-301 and SB-303 were built on single-sided circuit boards of different sizes with different layouts. A critical feature in both layouts was to place the audio summing amplifier in a location where it would be near the existing audio wiring in the receiver when the circuit board was installed. Adhesive copper tape was used on the back side for large size runs along with jumper wires for crossovers. No runs were cut on the top side of the circuit board so that it would be a continuous ground plane. Single integrated circuits were used to simplify construction for two-sided circuit card construction. All four integrated circuits are mounted in DIP sockets to allow jumper wire connections to the socket pins.

Control Logic

Two 2N7000 MOSFET switching transistors are used in the control logic to replace the obsolete Siliconix VMP-2 devices used in the SB-301 and SB-303 circuits. The first transistor interfaces with the external control signal and isolates the sidetone-disable control line. The two resistors on the gate form a voltage divider that bias the gate to +5 V dc to turn transistor Q1 on. The pull-up resistor divider allows TTL devices, CMOS devices, or open-collector devices running on +5 V dc supplies to control the sidetone. The internal pull-up resistors also allow the use of a hand key for code practice. The resistors provide the voltage to turn Q1 on, which disables the sidetone when the key is up. When the key is down, the gate voltage on Q1 drops to near zero, turning it off and enabling the sidetone.

The voltage divider on the gate of Q1 is designed for a positive supply voltage of +15 V dc. Scale the values of R1 and R2 if a different positive supply voltage is used to maintain a +5 V dc logic interface.

The second MOSFET provides a means of disabling the sidetone and controls the shaping circuit. When Q2 is off, it allows the shaping capacitor to charge to half the supply voltage. When it is turned on, it discharges the shaping capacitor to very close to zero volts. The disable control switch is connected directly to the gate of Q2, allowing it to override the keying input signal. Single-pole, single-throw (SPST) toggle switches were added to the SB-301 and the SB-303 to allow the sidetone to be enabled or disabled. The Q2 gate voltage divider sets the maximum gate voltage at +10 V dc for the lowest MOSFET on resistance. Reverse the values of R3 and R4 to match those of the input logic stage for external disable control with logic gates running on +5 V dc supplies in place of a toggle switch. The gate voltage dividers for Q2 are also designed for a positive supply voltage of +15 V dc. Scale the resistance values of R3 and R4 if a different positive supply voltage is used with a logic-controlled disable function.

Keying Shaping Circuit

A simple resistor-capacitor circuit is used to shape the voltage used to amplitude modulate the waveform generator output. The time constant for the voltage rise and fall is set by R7 and C2 at 2.3 ms. A stable polyester capacitor and stable metal film resistor are used in the shaping circuit to maintain constant rise and fall times. The time constant can be increased to soften the keying transitions or decreased to harden them by changing the values of R7 and C2. Resistors R5 and R6 should also be stable 1% tolerance metal film components of equal value. The voltage division ratio is important because it determines the maximum attenuation of the oscillator output signal in the key-up state. The XR2206 output amplitude decreases as the AM control voltage increases from zero and reaches a minimum −44 dB lower when the AM control voltage is half the supply voltage. The output amplitude

19 Ellison Rd
Lexington, MA 02421
Malcolm_Crawford@Raytheon.com

starts increasing with a reversal in the signal phase when the AM input voltage increases above the half supply voltage level.

An operational amplifier (U1) is used as a buffer amplifier to prevent the AM control input resistance of the waveform generator from loading down the shaping circuit's voltage divider. The waveform generator IC's load resistance would only have a small effect on the time constant but would lower the maximum sidetone attenuation by chang-ing the voltage division ratio. The unity gain operational amplifier uses a resistor (R27) in the feedback equal in value to the shaping resistor (R7) for the lowest output offset volt-age. Military temperature range LM101AH operational amplifiers were used in the two circuits that were built.

This op amp requires an external com-pensation capacitor (C3). A value of 100 pF was used to overcompensate the amplifier and increase stability. Industrial temperature range LM201A versions can also be substi-tuted for the LM301A version called out in the parts list. Other substitute operational amplifiers with the same pin outs, low offset voltage, and low noise performance are the TL071 and the TL081. Both are internally compensated and do not require the external compensation capacitor (C3) shown in the schematic. The LM741 operational amplifier is not recommended because of its unspeci-fied noise characteristics.
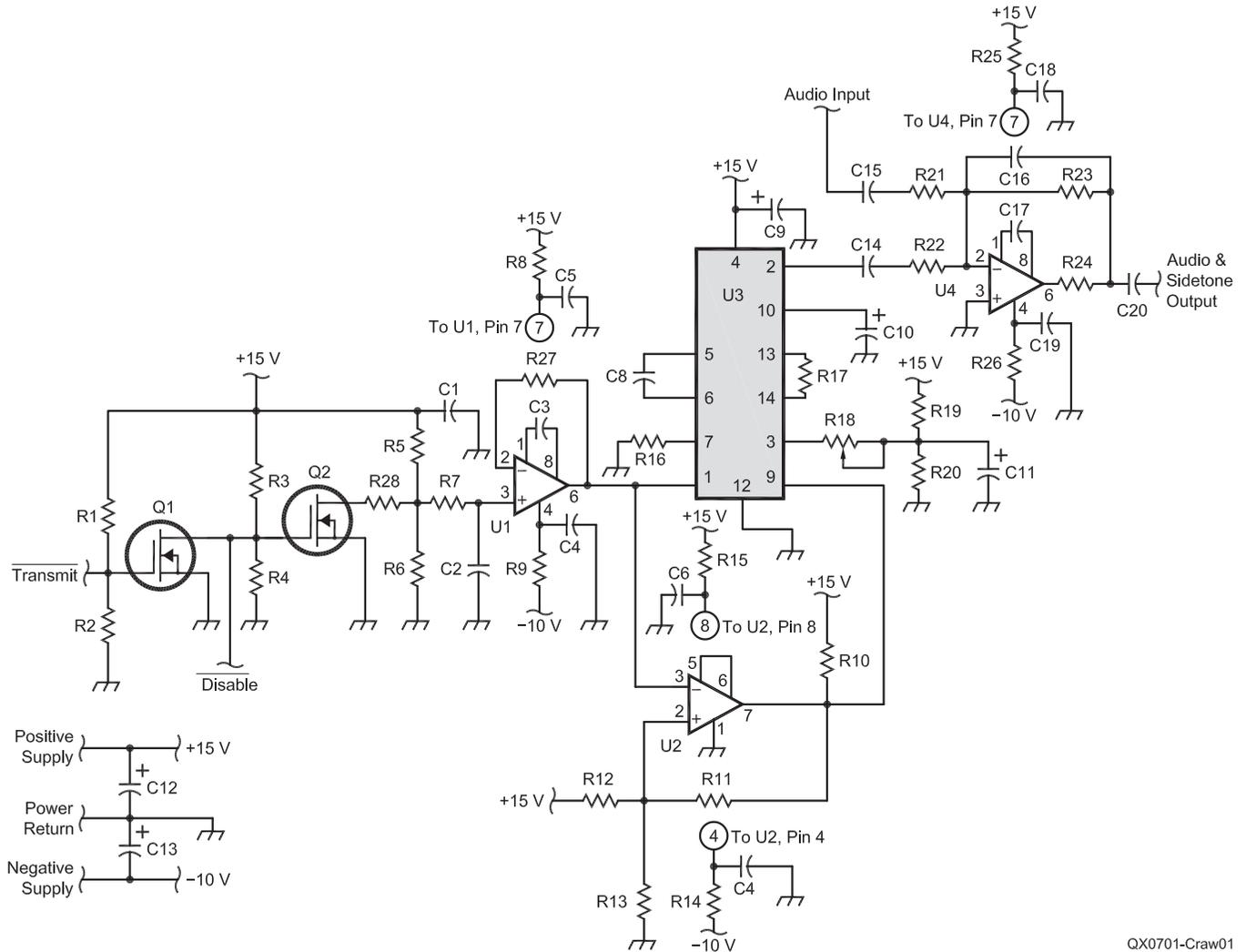
Figure 1—Schematic diagram of the Ultimate Sidetone. Integrated circuits and transistors were obtained from Jameco Electronics. An alternate source for the XR2206 is Newark Electronics. Alternate sources for all components except the XR2206 are Mouser and Digikey.

C1, C4, C5, C6, C7, C14, C18, C19 — 0.1 µF ceramic, 25 V
C2, C8 — 0.1 µF polyester 5%, 25 V
C3, C17 — 100 pF ceramic, 25 V
C9, C10 — 1 µF electrolytic, 25 V
C11, C12, C13 — 10 µF electrolytic, 25 V
C15, C20 — 0.1 µF ceramic, 25 V (SB-301), 1 µF ceramic 25 V (SB-303)
C16 — 47 pF ceramic 25 V (SB-301), 3600 pF ceramic 25 V (SB-303)
Q1, Q2 — 2N7000 MOSFET logic level transistor

R1, R4, R7, R10, R27 — 21.5 kΩ metal film 1% 0.1 W
R2, R3 — 10.7 kΩ, 5% 0.1 W
R5, R6, R13 — 2.49 kΩ metal film, 1% 0.1 W
R8, R9, R14, R15, R24, R25, R26 — 150 Ω, 10% 0.1 W
R11 — 402 kΩ, 10% 0.1 W
R12 — 2.61 kΩ metal film 1% 0.1 W
R16 — 12.4 kΩ metal film 1%, 0.1 W
R17 — 200 Ω 10% 0.1 W

R18 — 20.0 kΩ 24 turn ⅜ inch leaded trimmer potentiometer (Bourns 3296W-1-203, Spectrol 64W, Murata 3102W)
R19, R20 — 5.11 kΩ metal film 1% 0.1 W
R21, R23 — 316 kΩ metal film 1% 0.1 W (SB-301), 4.32 kΩ metal film 1%, 0.1 W (SB-303)
R22 — 316 kΩ metal film 1% 0.1 W
U1, U4 — LM301AN operational amplifier
U2 — LM311N voltage comparator
U3 — XR2206CP waveform generator

Figures 2 and 3 illustrate the key-up and key-down shaping envelope and the keyed and shaped tone, respectively.

Frequency Selection

An LM311 integrated circuit voltage comparator (U2) is used to select the frequency of the waveform generator. In the key-up state, the open-circuit timing resistor on pin 8 of the waveform generator is selected by a logic 0 at the comparator output to set the frequency below 0.01 Hz. The subaudible frequency is used to eliminate any residual sidetone after the maximum AM control attenuation level is reached. In the key-down state, the timing resistor on pin 7 (R16) is selected by a logic 1 on the comparator output to set the waveform generator frequency at 800 Hz. The pull-up resistor (R10) on the comparator output provides a logic 1 voltage well above the waveform generator pin 7 timing resistor select threshold

voltage of >+2 V dc. The typical comparator logic 0 output voltages of less than 150 mV dc are well below the pin-8 timing-resistor select threshold voltage of < +1 V dc.

Pins 5 and 6 of the comparator are connected together to slow it down to eliminate chatter on threshold crossings and oscillations. The comparator also has a hysteresis voltage of 43 mV to prevent the output from chattering when the input shaping voltage crosses the threshold. A good explanation of why comparators chatter and how hysteresis prevents it can be found in the *Hands on Radio* series installment on comparators.[1] Stable, 1%-tolerance metal film resistors are used in the threshold voltage divider circuit to provide an accurate threshold voltage slightly

[1]H.W. Silver, NØAX, "Hands on Radio Experiment #11 — Comparators," *QST*, Dec 2003, pp 55-56.

less than half the supply voltage. The threshold voltage was selected so that the 0.01 Hz to 800 Hz frequency switching occurs when the waveform generator output amplitude is near its minimum. Military temperature range LM111 versions and industrial temperature range LM211 versions of the comparator can be substituted for the LM311 version called out in the parts list.

Waveform Generator

The timing circuit of the waveform generator uses a stable, 1% metal film resistor (R16) and a stable, low-tolerance polyester capacitor (C8). The values in the parts list were selected to fall in the range that gives the lowest frequency sensitivity to ambient temperature changes. Use the formula…

$$frequency = 1 / (R16 \times C8) \qquad (Eq 1)$$

…to select different timing component values
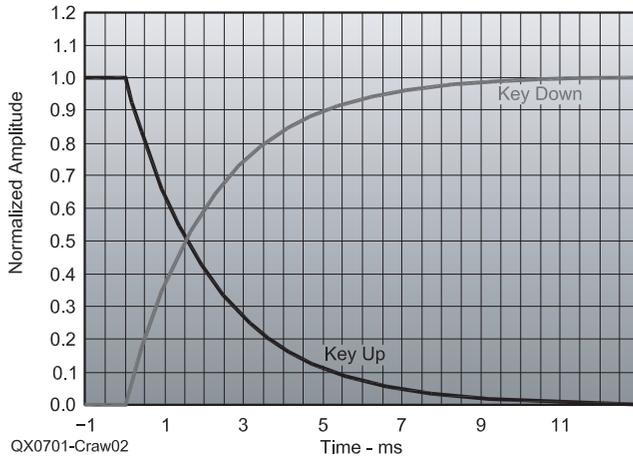
Figure 2—Key-up and key-down shaping envelope.
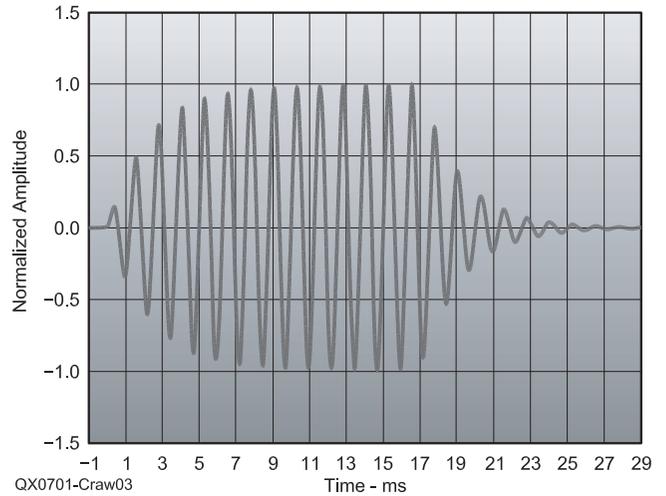
Figure 3—Keyed and shaped tone.
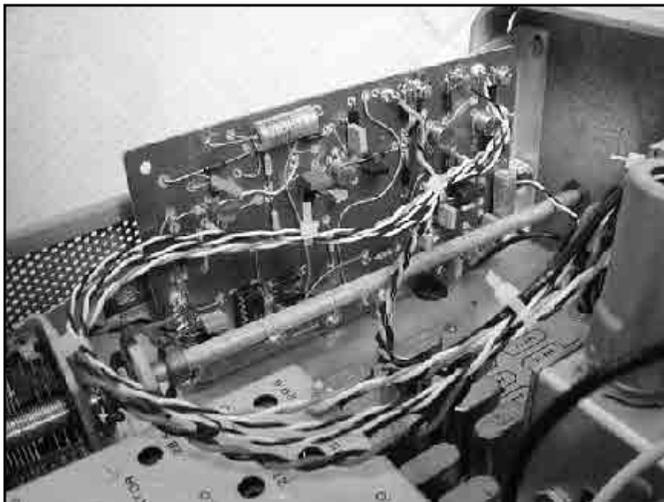
Figure 4— The installed SB-301 Ultimate Sidetone.
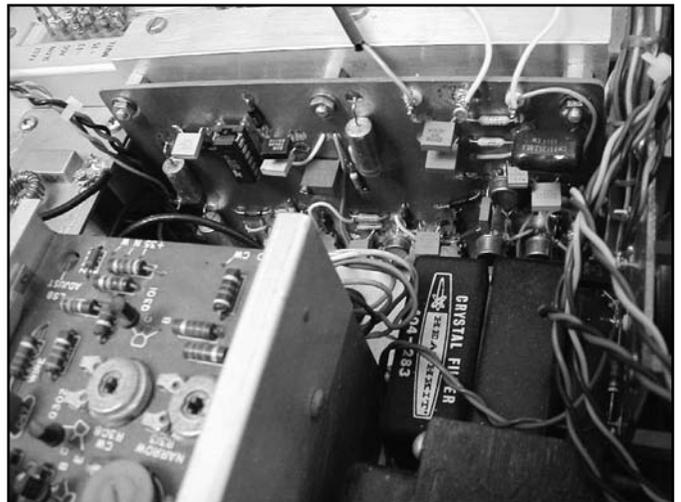
Figure 5— The installed SB-303 Ultimate Sidetone.
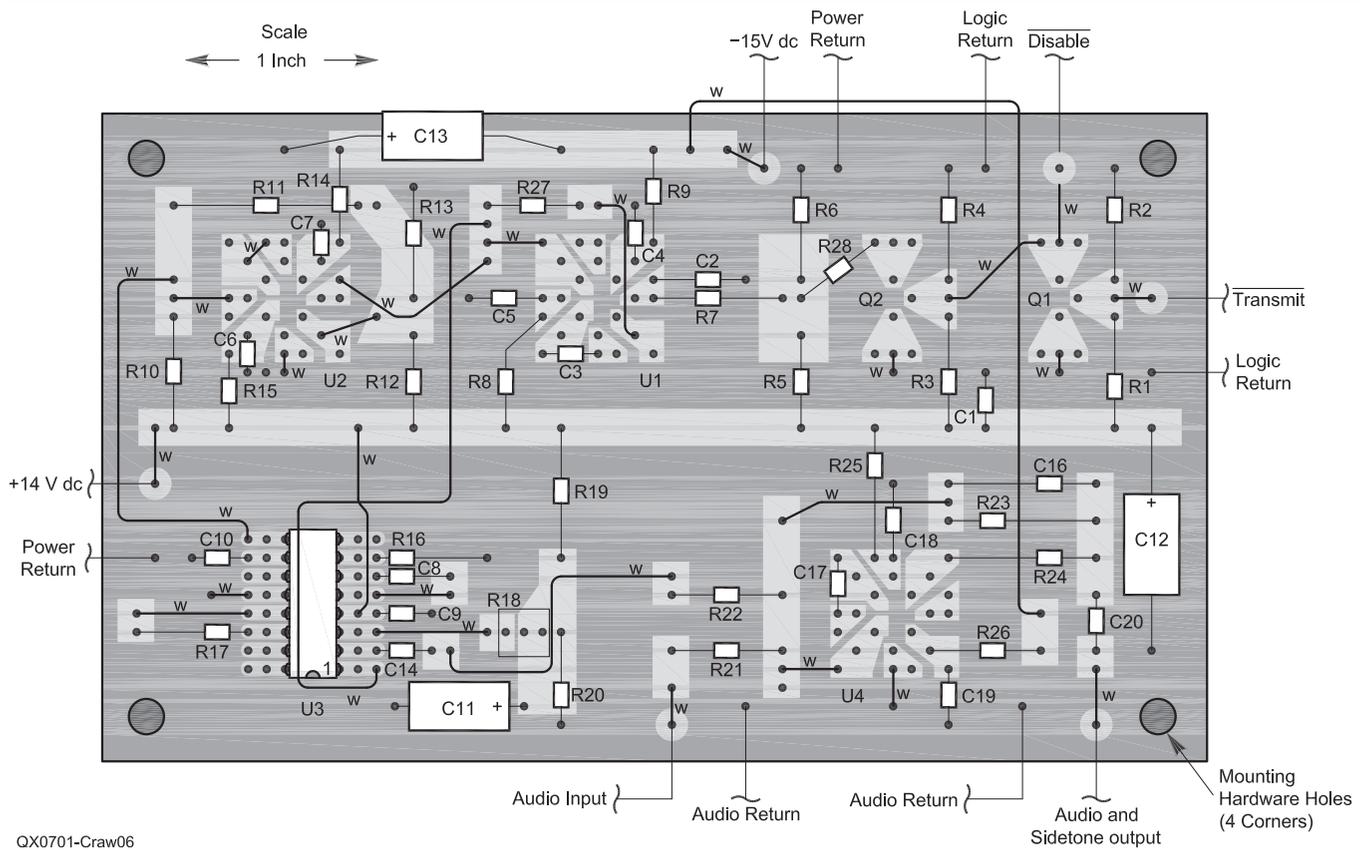
Figure 6— The circuit card layout for the SB-301 version.
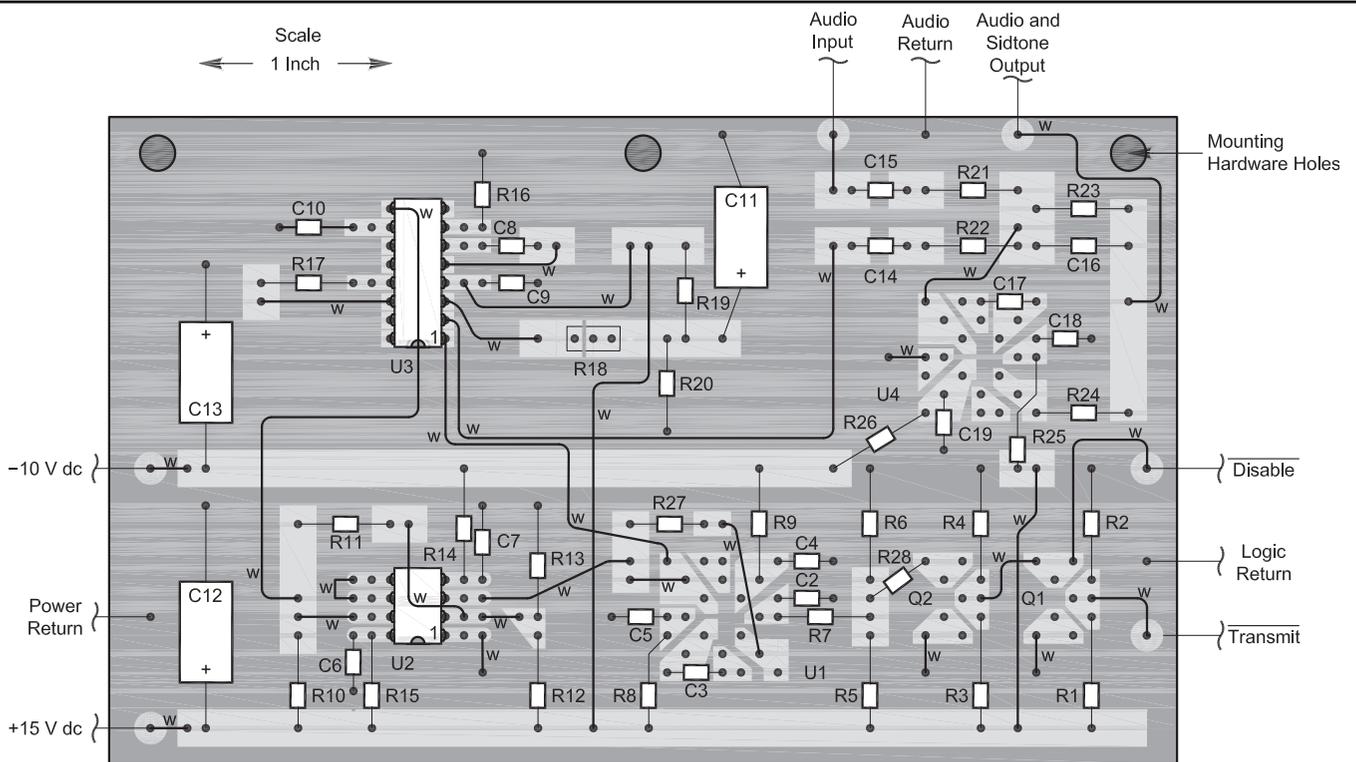
Figure 7— The circuit card layout for the SB-303.

to make use of available parts or to change the sidetone frequency from 800 Hz.

Stable 1%-tolerance metal film resistors (R19, R20) are used in the amplitude control circuit to set the sidetone amplitude control bias voltage at half the supply voltage. A multiturn potentiometer (R18) is used to provide a smooth adjustment for the sidetone amplitude. A capacitor (C14) is used to form a low-pass filter with R19 and R20 to attenuate any noise on the power supply that could modulate the waveform generator output amplitude. The XR2206C can operate on single supply voltages from +10 V dc to +26 V dc.

Summing Amplifier

A second op amp (U4) is used to combine the sidetone and receiver audio. The receiver audio high-pass corner frequency is set by R21 and C15 at 5.0 Hz for the SB-301 and 37 Hz for the SB-303. The low-pass corner frequency is set by R23 and C16 at 10 kHz for both receivers. The values of R21 and R23 are equal to achieve unity gain in the receiver audio path through the amplifier. Unity voltage gain is used for the receiver audio to avoid changing the output audio level. The different resistor values for the two receivers were selected to match the input resistance of first audio amplifier stages.

Setting the Ultimate Sidetone input resistance equal to that of the first audio amplifier stage was done to avoid changing the receiver's volume-control characteristic. Scale the values of C15 and C16 if you change the values of R21 and R23 to match the input resistance of your receiver. The receiver audio input high pass corner frequency should be less than 50 Hz and the low-pass corner frequency between 10 kHz and 20 kHz.

The sidetone audio high-pass corner frequency is set by R22 and C14 at 153 Hz for the SB-301 and 5.0 Hz for the SB-303. A dc-blocking capacitor is necessary on the output of the waveform generator because it sits at half the supply voltage. The voltage gain for the sidetone audio was set at unity for the SB-301 and at 0.014 for the SB-303 to reflect the different audio amplifier gains and headphone levels for the two receivers. If you use a different value for R22 scale the value of C14 to keep the high-pass corner frequency below 200 Hz.

Capacitors C15 and C20 may not be needed in some receivers that already have dc-blocking capacitors at the product detector output and the audio amplifier input. The Ultimate Sidetone used in the SB-301 deleted C15 because the product detector output contained a dc-blocking capacitor. The version used in the SB-303 deleted C20 because the audio amplifier input contained a dc-blocking capacitor.

A 100-Ω resistor (R24) is used inside the feedback loop of the summing amplifier to isolate the operational amplifier output stage from capacitive loads and provide additional short circuit protection. A compensation capacitor (C17) value of 100 pF was selected to overcompensate the LM301A op amp for increased stability. The alternate operational amplifiers in the keying shaping circuit description are also suitable for the summing amplifier stage by deleting the compensation capacitor (C17).

Power Requirements

The circuitry was designed to operate over a wide range of supply voltages. Supply voltages between +10 V dc and +15 V dc for the positive supply and –5 V dc and –15 V dc for the negative supply are suitable. Precision metal film resistors are used in the voltage divider circuits that determine critical operating voltages such as the keying shaping bias, comparator threshold voltage, and the waveform generator amplitude control voltage. The resistor dividers in these circuits scale the supply voltage to maintain the correct operating voltage levels without having to adjust the ratios for different supply voltages.

Positive and negative supplies were used to avoid half-supply biasing in the op-amp circuits. The two supply voltages need not be regulated for proper operation but should be well filtered. Large-value polarized capacitors (C12 and C13) are used on the positive and negative supply inputs to reduce audio noise levels and to improve the keying transient response of the circuits. The sidetone in the SB-303 uses the existing linear regulated +15 V dc and Zener-regulated –10 V dc supplies in the receiver. The estimated current drains are 33 mA for the +15V dc positive supply and 8 mA for the –10 V dc negative supply. The sidetone in the SB-301 uses unregulated +14 V dc and –15 V dc supplies that were added by the original owner (K9MO) for solid-state IF amplifier, AGC, and crystal-calibrator circuits. The two operational amplifiers and the voltage comparator have power supply decoupling resistors to improve the rejection of power supply noise above 10 kHz. Any resistor value between 100 Ω
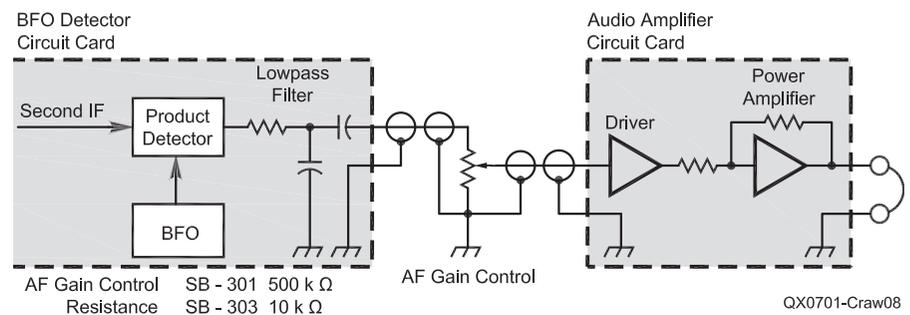
Figure 8—Block diagram of the unmodified receiver.
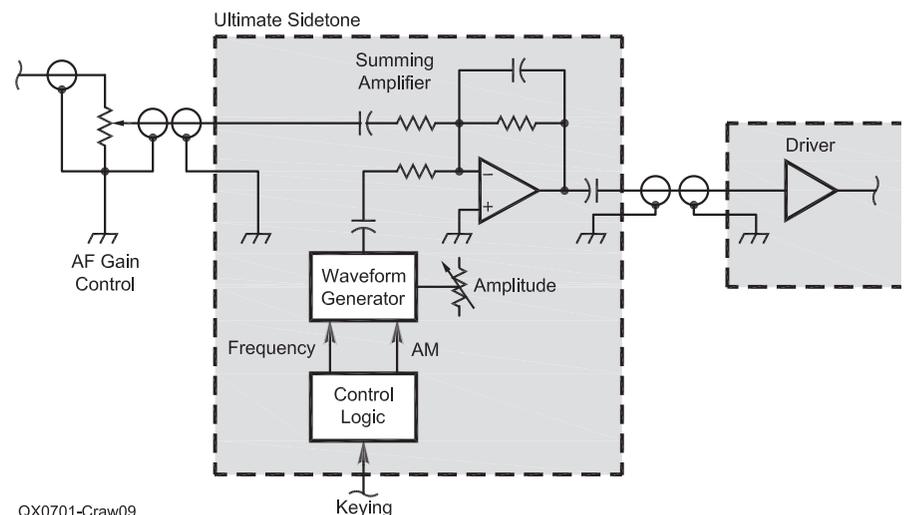
Figure 9—A block diagram of the modified receiver.

and 180 Ω can be substituted for the value in the parts list.

Building the Ultimate Sidetone

The sidetone circuit was built on a 3.5 × 6.0-inch single-sided, copper-clad board for the SB-301 and a 3.5 × 5.5-inch board for the SB-303. The layouts are very similar and the different sizes reflect the space available in the two receivers. The installed SB-301 Ultimate Sidetone is shown in Figure 4 and the installed SB-303 Ultimate Sidetone, in Figure 5. Teflon® insulated terminals are used for the audio input and output signal connections, logic input signal connections, and the power supply connections. Ground stud terminals are used for the audio, logic, and power return connections.

Another difference in the two layouts resulted from the use of a TO-99 metal case LM111 voltage comparator in the SB-301 instead of a dual in-line package (DIP) device. The layouts show the TO-99 metal case LM101A operational amplifiers used in both designs. For new construction I would recommend using DIP LM301A and LM311 integrated circuits and low profile sockets. The layout of the boards puts the summing amplifier close to the existing audio wiring in the two receivers to simplify the modifications needed for the Ultimate Sidetone audio input and output to connections. The circuit card layout for the SB-301 version is shown in Figure 6 and the SB-303 version in Figure 7.

The circuit cards were fabricated using self-adhesive copper tape for the power runs and interconnection pads on the unclad bottom side of the board. Reverse image photocopies of the layouts were made and glued to the unclad sides of the boards with rubber cement. The layout was then used as a guide for drilling holes and cutting the copper tape.

Receiver Modifications

Both receivers were modified by breaking the wiring between the AF gain control potentiometer and the input of the audio output amplifier stages. Figure 8 shows the block diagram of the original receiver audio circuitry and Figure 9 shows the modification needed for the Ultimate Sidetone. In both receivers, the front-panel AF gain potentiometer has shielded cable connections to the BFO circuit card and the audio amplifier circuit card. The shielded cable that connects to the AF gain potentiometer wiper arm is disconnected from the input of the audio amplifier circuit card and connected to the input of Ultimate Sidetone. A new shielded audio cable is run from the Ultimate Sidetone output to the audio amplifier circuit card input. Miniature coaxial cable can be used if shielded audio cable is not available. The overcompensation in the summing operational amplifier and the isolation resistor inside the feedback loop will keep the amplifier from oscillating with the capacitive load of the shielded cable.

The waveform generator output signal is much larger than needed in most receivers. The summing amplifier sidetone gain must be set to unity or lower as to reduce the sidetone signal amplitude to a comfortable listening level. The audio gain after the audio volume control in the SB-301 was ×200 and ×392 in the SB-303. The type of headphones used also makes a difference in setting the sidetone audio gain. The old high-impedance US Army Signal Corps headphones I use with the SB-301 require 514 mV RMS for a comfortable sidetone level. The lightweight low-impedance stereo headphones used with the SB-303 required only 81 mV RMS for a comfortable sidetone level. The values of R22 were selected so that the adjustment range of R18 could set the waveform generator output amplitude at a comfortable sidetone level.

The value of R22 for other receivers can be calculated by using the audio amplifier gain, the audio output level desired for the headphones or speaker used, and the output amplitude of the waveform generator. The output amplitude scale factor for the waveform generator is 42.4 mV RMS per kilohm of control resistance (R18). A minimum value of 1 kΩ for R18 is recommended when selecting a value for R22. A maximum value of 1MΩ for R22 is recommended to avoid having the parasitic leakage resistance of the circuit card setting the gain.

Mal Crawford, K1MC, was first licensed as a Novice in 1959 with the call sign WV2IPC. After earning BS and MS degrees in Electrical Engineering at Rensselaer Polytechnic Institute and completing active military service in the US Army Signal Corps, he has lived in New England and worked in the field of missile and radar electronics. Mal enjoys hiking and volunteer trail maintenance and construction activities when not designing, building, and operating home-built equipment. His recently completed HF transmitter encompasses a half century of electronic technology, ranging from a neutralized Class-C vacuum-tube amplifier to an integrated-circuit direct digital synthesizer. QEX

All About the Discone Antenna: Antenna of Mysterious Origin and Superb Broadband Performance

Learn about the development, history and some applications of a discone antenna.

Steve Stearns, K6OIK

"The frequency bandwidths demanded by high-definition television have considerable range…" With these prescient words, Philip S. Carter of RCA opened a 1939 paper that compared a variety of antennas for the emerging field of "high-definition" television. Carter showed conclusively that conical antennas held distinct advantages over dipoles and folded dipoles when it comes to broadband performance. Today, conical antennas are making a comeback for broadband applications such as digital television and UWB (ultra-wideband) or impulse radio. Stacked arrays of bowties and biconical dipoles are gradually displacing traditional mainstay antennas such as Yagis and log-periodics for the rooftop reception of digital television (DTV). One conical antenna, long popular among scanner hobbyists, the discone, has been described in previous articles in Amateur Radio magazines and books. The story has never been told fully, however. This article explains the history and theory of the discone, corrects some common misunderstandings, and presents an *EZNEC* model for a 0.6-octave discone that readers may copy and scale to their favorite frequency bands.

Conical antennas, and the discone in particular, have an obscure but fascinating history. Sergei Alexander Schelkunoff, at Bell Labs, was a titan of antenna theory in the early to mid 20th Century. In 1941, Schelkunoff published a major paper in the *Proceedings of the IRE*, which, among other things, analyzed the symmetric biconical dipole and showed that many other antennas can be analyzed

as extensions of it.[1] The discone antenna (Figure 1) is one such extension, in which the biconical dipole is asymmetric, one cone's angle being 90°, which gives a flat disk of radius equal to the cone length. Two years later, in 1943, Armig Kandoian at the Federal Telephone and Radio Corporation applied for a pat-

[1]Notes appear on page 43.

ent on the discone antenna. Kandoian's novel or inventive element was apparently that the antenna could be encased in a radome, making it suitable for aircraft, not that it used a cone or disk per se, those ideas being obvious in view of Schelkunoff's prior work. The patent was granted in 1945, whereupon Kandoian and his colleagues, Sichak, Felsenheld, and Nail, at the newly renamed Federal Telecommunica-
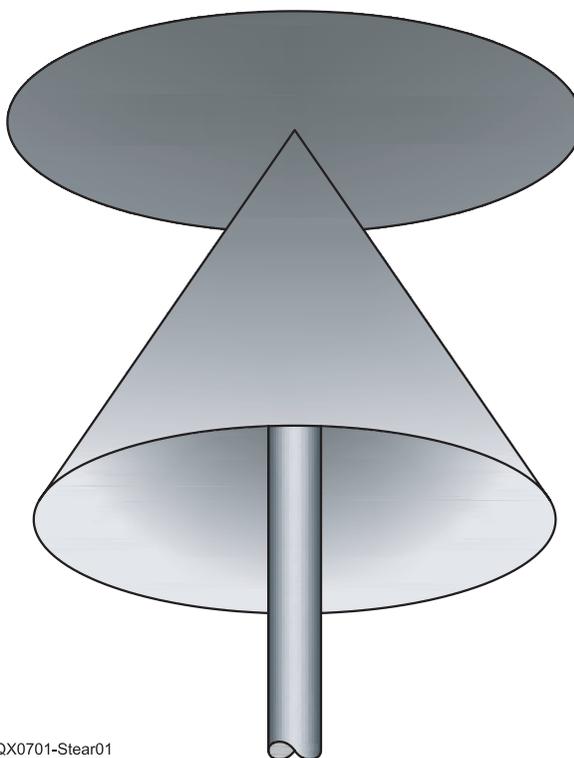
Figure 1 — This illustration shows a home-made discone for 2.4-GHz WiFi use. See www.spaziolink.com/wi-fi.

QX0701-Stear01

PO Box 4917
Mountain View, CA 94040-0917
k6oik@arrl.net

QEX Jan/Feb 2007 37

tion Laboratories, a subsidiary of ITT, began publicizing the antenna in a series of articles in various journals from 1946 to 1953.

In 1952, Schelkunoff published the book *Advanced Antenna Theory*, which gave a comprehensive analysis of the asymmetric biconical dipole in which the angles and lengths of both cones are arbitrary. The discone appeared on page 93 as a special case. Engineering studies of the discone followed shortly thereafter by Nail at the Federal Telecommunication Laboratories and by Crowley and Marsh at Ohio State University. Many variations on the basic discone have appeared since, having such features as multiple cones, multiple disks, meander lines for the cone, and mechanical tuning devices.

Radio Amateurs, meanwhile, had noticed this interesting antenna. A construction article appeared in *CQ* in 1949. More construction articles appeared since then, and are noted at the end of this article. Given such interest, it is surprising that amateur antenna modelers have largely overlooked this antenna. This article corrects that oversight by presenting an *EZNEC* model for a discone that readers may copy, modify, or scale to their favorite bands.

Conical antennas consisting of a single cone fed at its apex against an infinite ground plane are often called "monocones" or less often "unipoles." If the infinite ground plane is replaced by one that is finite and circular, the antenna is called a "discone." A discone can also be thought of as an asymmetric biconical dipole in which one cone's angle is 90° (measured from its axis), so it opens to become a flat disk. The impedance of a discone depends on frequency and three geometric variables: the cone's angle, slant length (measured along the

side of the cone), and the radius of the ground plane disk. Feed line SWR depends additionally on the line's characteristic impedance. A discone is not a frequency independent antenna, although this is a common misconception. Rather, a discone behaves more like a fat dipole. Its feed point resistance and reactance vary with frequency, although not through the extremes of a dipole.

Discones are used for broadband operation at frequencies above their first resonance. Manufacturer's data for two popular VHF/UHF discones, the AOR DA3000 and RadioShack 20-043 are shown in Figures 2 and 3 as graphs of return loss versus frequency. The vertical scale of the AOR curve is 10 dB/division; the scale of the RadioShack curve is unspecified. The key feature is that the curves are scalloped. The SWR cycles between high and low as frequency is varied. Receiving is possible on any frequency, but transmitting is best done in the SWR valleys. A good design will keep the SWR peaks below a design limit and position the valleys to coincide with desired transmit frequencies.

There are, broadly speaking, two methods for analyzing antennas that don't require construction and measurement. The first method is mathematical analysis, and the second is numerical antenna modeling. The former was the only method available before computers were invented. Antennas were analyzed mathematically by "normal mode theory" or by solving integral equations. In this article, we'll use a formula developed by Schelkunoff for the feed point impedance of a finite cone over an infinite ground plane, derived from spherical mode theory. The formula allows us to quickly determine the best length and angle

for a cone depending on design impedance and bandwidth. More exact formulas for when the disk radius is finite are in the engineering literature. We'll use an *EZNEC* model when analyzing such cases.

In his 1941 paper, Schelkunoff showed that the feed point impedance of many antennas, including the conical monopole over a ground plane, can be represented as terminated transmission lines one-quarter wave shorter than the length of the antenna:

$$Z_{in} = Z_0 \frac{Z_m + jZ_0 \tan\left(k\ell - \frac{\pi}{2}\right)}{Z_0 + jZ_m \tan\left(k\ell - \frac{\pi}{2}\right)}$$
(Eq 1)

For a conical monopole of angle θ, measured from axis, the characteristic impedance Z_0 is given by:

$$Z_0 = \frac{\eta}{2\pi} \ln \cot\left(\frac{\theta}{2}\right)$$
(Eq 2)

The terminating impedance $Z_m = R_m + jX_m$ is the radiation impedance referenced to the current maximum on the antenna. Schelkunoff gave general formulas for the real and imaginary parts of Z_m for all cone angles, but he also gave the formulas for small cone angles in Equation 3 below, where k is the wavenumber $2\pi/\lambda$ and η is the characteristic impedance of free space equal to $\mu_o c$, the speed of light times the magnetic permeability of free space, or 376.73 Ω. (It would be exactly 120π Ω if light would cooperate and travel at exactly 300 million meters per second.) Other symbols in the formulas are Euler's constant, $C = 0.5772156649\ldots$ and

$$R_m = \frac{\eta}{4\pi}\left\{C + \ln(2k\ell) - \text{Ci}(2k\ell) + \frac{1}{2}\sin(2k\ell)\left[\text{Si}(4k\ell) - 2\text{Si}(2k\ell)\right] + \frac{1}{2}\cos(2k\ell)\left[C + \ln(k\ell) + \text{Ci}(4k\ell) - 2\text{Ci}(2k\ell)\right]\right\}$$

(Eq 3)

$$X_m = \frac{\eta}{4\pi}\left\{\text{Si}(2k\ell) - \frac{1}{2}\sin(2k\ell)\left[C + \ln(k\ell) - \text{Ci}(4k\ell)\right] - \frac{1}{2}\cos(2k\ell)\text{Si}(4k\ell)\right\}$$

QX0701-Stear02

Start 25.000 000 MHz   Stop 2 600.000 000 MHz
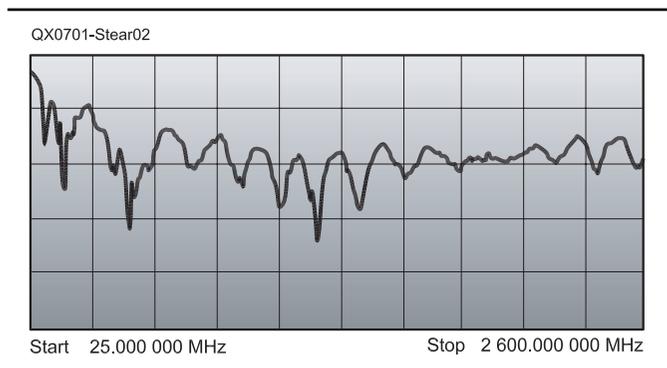
Figure 2 — Return loss of AOR DA3000 discone antenna.

Frequency (MHz)

25   500   955   1300

QX0701-Stear03
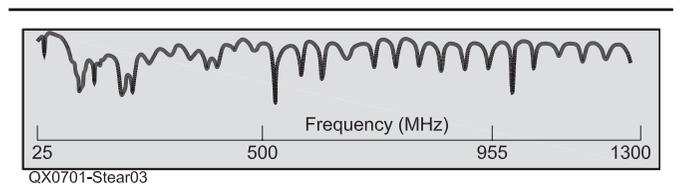
Figure 3 — Return loss of RadioShack 20-043 discone antenna.
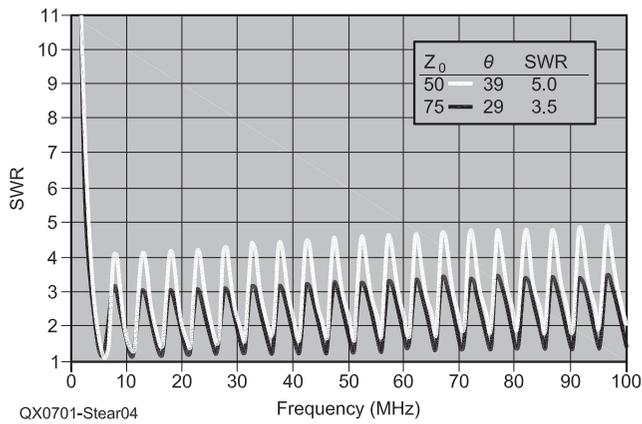
Figure 4 — Computed SWR (at 50 and 75 Ω) of two 100-foot cones.

Z₀	θ	SWR
50	39	5.0
75	29	3.5
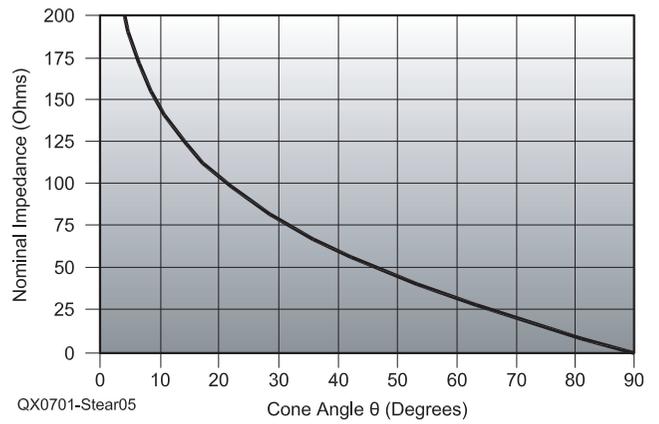
Figure 5 — Nominal monocone and discone impedance versus cone angle.
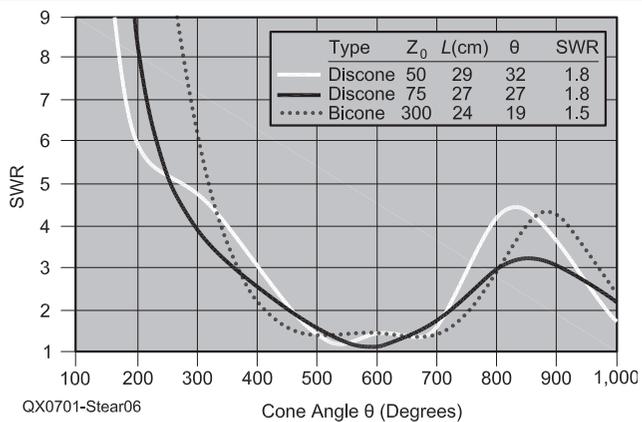
Figure 6 — Predicted SWR of three antennas optimized for UHF TV.

Type	Z₀	L(cm)	θ	SWR
Discone	50	29	32	1.8
Discone	75	27	27	1.8
Bicone	300	24	19	1.5

the sine and cosine-integral functions $Si(x)$ and $Ci(x)$, which we won't explain here.

Schelkunoff's asymptotic formula isn't numerically accurate for discones having large cone angles or finite disk radii. The formula, however, does reveal general trends and interesting design trades. More accurate formulas for general discones and biconical dipoles or "bicones" were developed by Hahn and Fikioris, and most recently by Samaddar and Mokole.[2]

For broadband operation, the best cone angle depends on bandwidth. Given a frequency band from f_1 to f_2, the optimum cone angle decreases as the ratio f_2/f_1 increases. For a nominal 50-Ω antenna, as the design bandwidth increases from one to five octaves, the optimum cone angle decreases from 47° to 39°, and the peak SWR creeps up. In addition, as the feed point design impedance increases, the optimum cone angle decreases. An interesting implication pursued in some designs is that the cone should be curved instead of flat sided. Our interest here is flat-sided cones.

Figure 7 — Wire table of an *EZNEC* discone model.

Figure 4 shows SWR curves calculated from Schelkunoff's formula for two 100-foot cones over infinite ground planes. The cone angle and length have been optimized for five-octave operation at both 50 and 75 Ω. Now, we can use an antenna modeling program to get better accuracy with less effort. Nonetheless, I found the best cone angles for each feed point impedance in a five-octave band with a single command to Microsoft *Excel*'s solver tool and 60 seconds of patience. Doing the same optimization in *EZNEC* would have taken days if *EZNEC* had an optimizer — which it doesn't, unfortunately.

An approximate formula for the best cone angle for a given feed point design impedance is obtained from the characteristic impedance formula above.

$$\theta = 2\tan^{-1}\exp\left(\frac{-Z_0}{60}\right)$$

(Eq 4)

Figure 5 illustrates the relation between θ and Z_0 given by this approximation. The predicted angle is good for design bandwidths up to two octaves but should be reduced if the design bandwidth is greater.

There are a lot of different ideas about the proper shape of a discone. Typing "discone" into Google Images reveals a variety of shapes. A common error appears to be making the disk too small and the cone too long. Using a computer, one can jointly optimize a cone's angle, slant length, and disk radius. Increasing the disk radius while simultaneously decreasing the cone's slant length is akin to sliding a feed point along an off-center-fed (OCF) dipole. This interpretation becomes exact if we regard the discone as an OCF biconical dipole with one cone's angle being 90°. Computer modeling reveals the best geometry for a given design impedance and band of operation, as will be shown below.

The procedure for designing a discone for transmitting has one extra step. The slant length is adjusted to put the SWR valleys on the desired transmit frequencies. Alternatively, an SWR valley can be shifted to a transmit frequency by using a mechanical tuning scheme such as those of McNamara or Rappaport.

When constructing a discone, the cone and ground plane can be made from rods or sheet metal as illustrated in Figure 1. When using rods, at least eight should be used. The AOR DA3000 uses 16, while the Diamond D-130J and RadioShack 20-043 use eight. You can adjust the impedance by bending the rods in or out. This is an advantage of rod construction.

Example

As an example, we'll consider a discone for receiving UHF TV channels 14 through 53. The frequency range is 470 MHz to 710 MHz. We set the discone's first resonance at a frequency below 470 MHz because, as shown in Figure 4, the SWR shoots up below the first resonance. Making the antenna too small incurs a big penalty.

A rule of thumb is to set the first resonant frequency at 0.7 times the lowest operating frequency. In this example, that comes out to 329 MHz or a wavelength of 91 cm. The disk radius plus cone slant length should equal half of this number or 46 cm. Now, you could allocate this length equally to the disk radius and cone slant length, making them both 23 cm. This may not be the best way to divide the length, however. Nail suggests that for a 50-Ω design, the ratio of radius to cone length should be:

$$\frac{R}{L} = 0.72 \times \sin\theta$$

(Eq 5)

which gives $R/L = 0.36$ or the ratio $R{:}L = 26{:}74$ for a cone angle of θ = 30°. An antenna modeling program can be used to confirm this ratio or to find a better ratio for a different design impedance. You can vary the proportions: 10:90, 20:80, 30:70, 40:60, 50:50 and so on, and compute an SWR sweep for each combination to find what ratio gives the smallest peak SWR over the band of interest.

Let's consider the UHF TV antenna example. To keep things simple, we'll let the disk be an infinite ground plane and use Schelkunoff's asymptotic formula; in practice, we'd use *EZNEC* and include disk
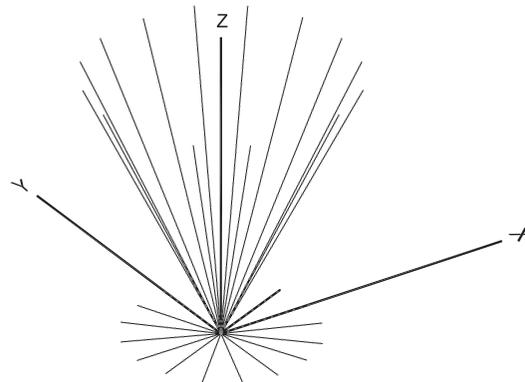
Figure 8 — Geometry of *EZNEC* discone model.

QX0701-Stear08

QX0701-Stear09
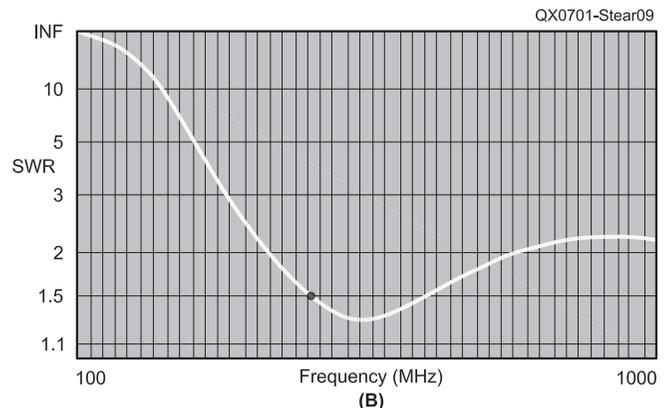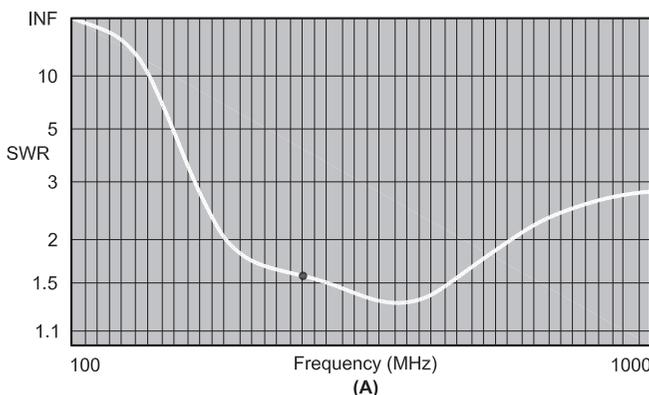
Figure 9 — SWR sweeps at (A) 50 Ω, and (B) 75 Ω; the marker is at 460 MHz.

radius as a variable. The goal is to find the cone angle and slant length that together minimize the maximum SWR between 470 MHz and 710 MHz. We'll find design dimensions for 50 and 75 Ω discones and a 300 Ω biconical dipole. A biconical dipole has a balanced feed point. When designing television antennas having balanced feed points, it is customary to make the feed point 300 Ω because this permits using both 300-Ω balanced twin lead and 75-Ω coaxial line with a 4:1 current balun.

Numerical optimization quickly finds the best lengths and cone angles. The optimum lengths are 29, 27, and 24 cm, with cone angles of 32°, 27°, and 19° respectively for 50 and 75-Ω discones and a 300-Ω bicone. Notice that the lengths are greater than those given by the rule of thumb. The reason is that the design bandwidth is narrow enough that lengthening the antenna moves an SWR valley down to fit the band. A more revealing explanation will be given shortly on a Smith Chart. Figure 6 shows the predicted results.

For a 75-Ω design impedance, the best combination of cone angle and length were found to be 27° and 27 cm. The predicted SWR is the darker curve in Figure 6. The maximum SWR between 470 and 710 MHz is predicted to be 1.82.

At this point we are ready to consider the effect of a finite disk radius. We'll check the theoretical predictions by using an *EZNEC* antenna model that includes a finite disk having Nail's recommended radius.

$$R = L \times 0.72 \times \sin \theta \qquad \text{(Eq 6)}$$
$$R = 27 \times 0.72 \times \sin 27° = 27 \times 0.72 \times 0.4540$$
$$= 8.8 \text{ cm}$$

The discone model's wire table and geometry are shown in Figures 7 and 8. The model was created easily by defining two wires as prototypes for the disk and cone, and then

using *EZNEC*'s radial tool to complete the model. Wire 1 is the prototype wire for the disk. Wires 2 through 16 were created by *EZNEC*. Similarly, Wire 17 is the prototype wire for the cone, and Wires 18 through 32 were created by *EZNEC*. It's convenient to think of cone and disk wires as being grouped into 16 pairs, with 31 segments allocated to each pair. Segment lengths are made nearly equal by allocating 25% of the segments to the disk and 75% to the cone. Disk wires, therefore, have 8 segments, and cone wires have 23 segments. This gives segment lengths of 11 mm for the disk and 11.7 mm for the cone wires. The apex of the cone was offset by 11 mm from the plane of the disk to make room for a single-segment source wire, which is Wire 33. The total number of segments in the model is 497, and the segment size is under $\lambda/25$ up to 1 GHz, which is well above the upper band limit of 710 MHz.

Figure 9 shows the SWR predicted by *EZNEC* for 50 and 75-Ω reference impedances. The graphs' vertical scales are nonlinear in SWR but linear in reflection coefficient magnitude. Allowing for graph distortion created by the nonlinear scale, the 75-Ω SWR curve on the right can be compared to that for the 75-Ω conical monopole shown in Figure 6 (darker curve) which was computed from Schelkunoff's asymptotic formula. The two curves are highly similar in both shape and value. It's clear that the dimensions obtained by optimizing Schelkunoff's formula are quite good, but there's still room for improvement. At this point one might choose to either build and test the antenna with the current dimensions or refine the *EZNEC* model.

It is noted that the model performs well as a 50-Ω antenna as shown in the left curve of Figure 9A. The computed SWR is 1.54 and 1.58 at the band edges and achieves a minimum of 1.30 at 610 MHz. Although not

explicitly optimized for 50 Ω, the dimensions are fairly good for that impedance too. This is not mere coincidence but a consequence using Nail's recommended disk size, which is for a 50-Ω design, rather than a 75-Ω design.

EZNEC's 500-segment restriction limits the bandwidth for which it can be used. A minimal NEC model would have eight wires for the cone and eight wires for the disk. If the length of each wire is a quarter wavelength or $\lambda_1/4$ at the lowest frequency f_1, then the total length of all 16 wires is $8\lambda_1$. The segment length should be no greater than $\lambda_2/20$ where f_2 is the highest frequency. The number of segments, obtained by dividing the segment length into the total length, is $80\lambda_1/\lambda_2$. Because *EZNEC* can handle at most 500 segments, the frequency ratio cannot exceed $f_2/f_1 = 500/80 = 6.25$, or 2.6 octaves. So, very broadband design should be done with a modeling program that can handle more than 500 segments, at least $80 f_2/f_1$ segments.

It's always a good idea to check whether a simple impedance matching network can improve the match over the band. The first step when designing a matching network is to plot the antenna impedance data on a Smith Chart. We'll use the impedance data that *EZNEC* computed. *EZNEC*'s frequency sweep feature allows the option of creating output data files for *MicroSmith* or *winSMITH*. It's best to choose *MicroSmith* to avoid *winSMITH*'s limit to 15 frequencies. *EZNEC* puts complex reflection coefficient (scattering parameter S_{11}) data in a .GAM text file. It should be opened with Microsoft Word, where it can be manipulated into a standard format for whatever EDA program you use, such as *ARRL Radio Designer*, Ansoft *Serenade SV* (featured in January 2001 *QST*), Agilent *ADS*, AWR *Microwave Office*, *RFSim99*, or even good old *SPICE*. I have found that Ansoft *Serenade SV* has the best capabilities for the money.
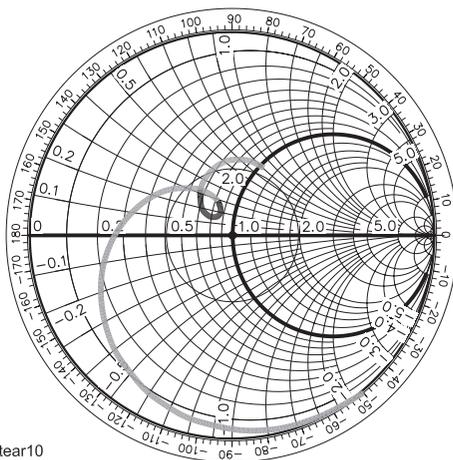
Figure 10 — Discone impedance on 75 Ω Smith Chart; UHF TV band highlighted with a darker band.
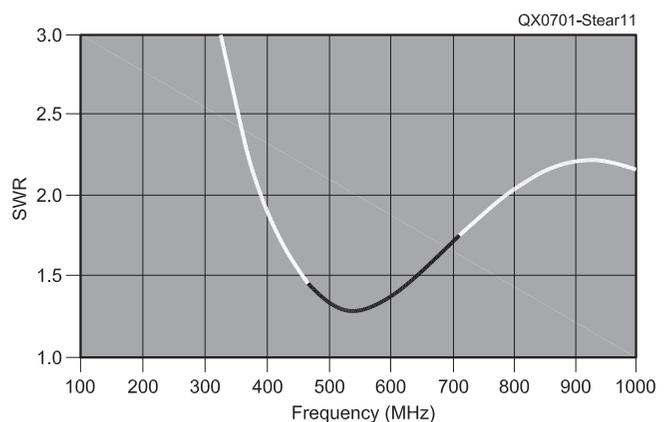
Figure 11 —SWR of unmatched discone (referenced to 75 Ω).
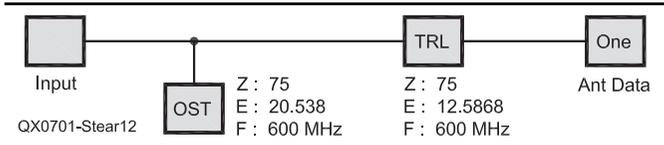
Figure 12 — An open stub impedance-matching network for 75 Ω.

Figure 13 — The match performance of an open stub is shown on a 75 Ω Smith Chart.
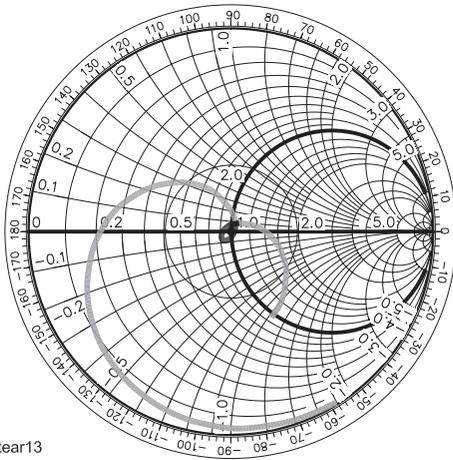
Figure 14 — The matching stub reduces the maximum SWR from 1.28 to 1.12 on the UHF TV band.

Table 1
An *EZNEC* .GAM Date Table Converted to .FLP Format for Analysis by *Serenade*

```
antdata 100MHz 1000MHz 91 50 S
EZNEC data for UHF TV discone antenna created on 2/28/2006.
  100MHz   0.9957904  -34.86269
  110MHz   0.9935989  -38.95737
  120MHz   0.9905544  -43.24265
                 •
                 •
                 •
  980MHz   0.4676752   42.40280
  990MHz   0.4689465   41.08628
 1000MHz   0.4697061   39.79027
```

The .GAM file is formatted to Ansoft's .FLP format within Microsoft *Word* in a few simple steps. First, delete the header line, leaving only the data lines. Next, use the text-to-table converter in *Word* to put the data into a four column table. Cut the contents of columns two and three and paste to columns three and four, leaving column two empty. Type the frequency unit "MHz" as the first entry in Column two, and paste it into all cells down the column. Next, do a table-to-text conversion, specifying a "space" character as the delimiter. Finally, remove the space between the frequency number and its unit by a global replacement of "[space]M" with "M." The data lines are now finished. Just add two header lines before the data lines, making sure to specify "50 S" on the first line to indicate that the data is scattering parameter data referenced to 50 Ω. This is the same convention

that *EZNEC* used when making the output data. Finally, save the file as a text file with the .FLP name extension to a *Serenade* project folder. The file should look like Table 1

Once the .FLP file has been saved, we open *Serenade SV* and define a one-port that references the file to represent the discone antenna. Run a frequency sweep, then use the report editor in *Serenade SV* to graph the antenna impedance on a Smith Chart by asking for a polar plot of S_{11} and specify Z or Y coordinates, or both.

The discone model's impedance is presented in Figures 10 and 11. Both figures assume a 75 Ω reference impedance. Figure 10 shows the complex impedance curve on the Smith Chart. For antennas, which are passive loads, the curve bends clockwise as frequency increases. Figure 11 shows the resulting SWR, whose agreement with
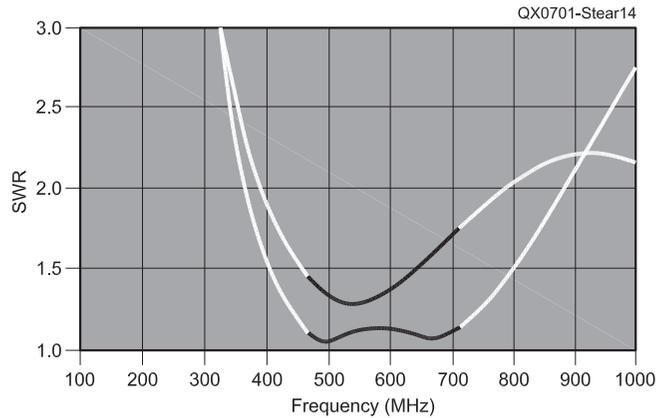
Figure 9B confirms that the *EZNEC* data conversion was done correctly. The data is plotted for the one-decade band from 100 MHz to 1,000 MHz. The 0.6-octave UHF TV band, from 470 MHz to 710 MHz, is highlighted with a darker line. This is the region where we want to match the impedance.

The secret to understanding the behavior of this medium bandwidth discone is to note that discone impedance curves aren't uniform concentric spirals on a Smith Chart, like a dipole would be. Instead, the impedance curves have a small loop in the middle of every large loop. By varying the cone angle and disk radius, a small loop can be moved to the center of the Smith Chart. Then, by merely scaling the dimensions, nearly an octave of bandwidth can be slid into the small loop. This yields a moderately broadband low-SWR antenna, such as our example discone.

These steps can be done in reverse order: first put the UHF TV band in a small loop and then move the loop to the center of the Smith Chart by inserting an impedance matching network at the discone's feed point. With more work, however, the steps can be done in the original order, and the impedance matching network eliminated.

The simplest impedance matching network for the discone model is the 75 Ω open stub shown Figure 12. The stub was designed to match the antenna to 75 Ω and is made of 75-Ω transmission line. The stub is inserted in a 75-Ω feed line at the proper distance from the feed point. The stub's electrical length and position from the feed point (20.5° and 12.6° at 600 MHz) translate to physical lengths of 29 mm and 17 mm times the velocity factor of the transmission line. The network can be constructed by using a 75-Ω coaxial T connector. Because a physical stub terminates in a fringing capacitance rather than an ideal

infinite-impedance open circuit, a real stub must be made shorter to achieve the predicted performance. Rather than calculate the fringing capacitance, it's easy to trim the stub by measurement during construction.

The impedance matching performance of the stub matching network is shown in Figures 13 and 14. Figure 14 shows before and after SWR curves. The matching network reduces the maximum SWR in the UHF TV band from 1.28 to 1.12.

Although most amateurs think of low SWR as important for transmitting, it is also important for receiving digital modulations such as DTV signals. The game here is not about power transmission. Rather, it is about avoiding waveform distortion caused by frequency selectivity of the communications channel. Wideband digital signals hate reflections, regardless of source. Reflections from multipath propagation and transmission line discontinuities are equally bad. The question of where reflected power goes is, ultimately, unimportant because communication is about getting information through, not power. Power transmission is merely a means to a greater end, not the end in itself. Reflections should be avoided.

Discones and bicones are better antennas for receiving HDTV signals than bow-tie or flat triangle antennas although the latter are better than log periodics and Yagis when phase distortion is considered. A bicone is easier to design than a discone because you build two identical cones. The question of disk size disappears. That's one fewer variables to get right. Other things to consider are pattern and polarization. Antennas should be mounted with the correct polarization — vertical for VHF/UHF communication signals and horizontal for receiving FM and television broadcast signals. When mounted horizontally, the azimuthal gain pattern is like that of a horizontal dipole — a figure eight for low frequencies and increasingly multi-lobed as frequency increases. At high frequencies, a discone's main lobe lies in the half-space on the cone side of the disk. As frequency increases from low to high, the main lobe shifts from the plane of the disk toward the direction of the cone, and minor lobes emerge on both sides of the disk.

Acknowledgments:

Parts of this article appeared in the newsletter of the Foothills Amateur Radio Society, an ARRL affiliated club. Those portions are included here with permission.

Notes
[1] The Institute of Radio Engineers (1912-1963) merged with the older American Institute of Electrical Engineers (1884-1963) to form the Institute of Electrical and Electronics Engineers (IEEE) in 1963. The IRE was instrumental in the creation of the Federal Radio Commission in 1927, which became the FCC in 1934.

[2] Mathematically skilled readers will find the papers by Hahn and Fikioris and by Samaddar and Mokole contain rigorous extensions of Schelkunoff's original analysis.

Steve Stearns, K6OIK, started in Amateur Radio while in high school at the height of the Heathkit era. He holds an FCC Amateur Extra license and a commercial General Radiotelephone Operator license with Radar endorsement. He previously held Novice, Technician, and 1st Class Radiotelephone licenses. He studied electrical engineering at California State University Fullerton, the University of Southern California, and Stanford University, specializing in electromagnetics, communication engineering and signal processing. Steve works at Northrop Grumman's Electromagnetic Systems Laboratory in San Jose, California, where he developed smart antenna and digital demodulation technology for eliminating interference and is currently developing ways to design and fabricate metamaterials to control electromagnetic phenomena. He has over 50 professional publications and eight US patents and numerous foreign patents. Steve is an ARRL Life member, has served as an assistant director of ARRL Pacific Division, and is currently the vice-president of Foothills Amateur Radio Society serving Silicon Valley. Steve has received numerous awards for professional and community volunteer activities. He can be reached at **k6oik@arrl.net** *or* **stearns@ieee.org***.*

Bibliography
Engineering literature:
Howe, G. W. O., "The Nature of Electromagnetic Waves Employed in Radio-Telegraphy and the Mode of Their Propagation," *The Electrician*, 19 Sep 1913, pp 965-969.

Howe, G. W. O., "The Capacity of an Inverted Cone and the Distribution of its Charge," *Proceedings of the Physical Society of London*, Vol 29, 1917, pp 339-344.

Shelkunoff, S. A., "Transmission Theory of Spherical Waves," *AIEE Transactions*, Vol 57, 1938, pp 744-750.

Carter, P. S., "Simple Television Antennas," presented at National Convention of IRE, San Francisco, Jun 1939. Published in *RCA Review*, Vol 9, Oct 1939, pp 168-185.

Schelkunoff, S. A., "Theory of Antennas of Arbitrary Size and Shape," *Proceedings of the IRE*, Vol 29, Sep 1941, pp 493-521; corrections published in Nov 1941, p 603, and Jan 1943, p 38. Reprinted without corrections in *Proceedings of the IEEE*, Vol 72, no. 9, Sep 1984, pp 1165-1190. Refer to Figure 2A with θ_1 or θ_2 equal to $\pi/2$.

Kandoian, A. G., *Broad Band Antenna*, U.S. patent no. 2,368,663, 6 Feb 1945.

Kandoian, A. G., "Three New Antenna Types and Their Applications," *Waves and Electrons*, Vol 1, no. 2, , Feb. 1946, pp 70W-75W, a separate section within *Proceedings of the IRE*, Vol 34. Reprinted in *Electrical Communication*, Vol 23, no. 1, Mar 1946, pp 27-34.

Kandoian, A. G., W. Sichak, and G. A. Felsenheld, "High Gain with Discone Antennas," *Proceedings of the National Electronics Conference*, Vol 3, Chicago, Nov. 1947, pp 318-328. Reprinted with additional figures in *Electrical Communication*, Vol 25, no. 2, June 1948, pp 139-147.

Smith, P. D. P., "The Conical Dipole of Wide Angle," *Journal of Applied Physics*, Vol 19, no. 1, Jan 1948, pp 11-23.

Tai, C. T., "On the Theory of Biconical Antennas," *Journal of Applied Physics*, Vol 19, no. 12, Dec 1948, pp 1155-1160.

Tai, C. T., "Application of a Variational Principle to Biconical Antennas," *Journal of Applied Physics*, Vol 20, no. 11, Nov 1949, pp 1076-1084.

Leitner, A. and R. D. Spence, "Effect of a Circular Groundplane on Antenna Radiation," *Journal of Applied Physics*, Vol 21, no. 10, Oct 1950, pp 1001-1006.

Schelkunoff, S. A., "General Theory of Symmetric Biconical Antennas," *Journal of Applied Physics*, Vol 22, no. 11, Nov 1951, pp 1330-1332.

Schelkunoff, S. A., *Advanced Antenna Theory*, Section 2.21, Wiley, 1952. Refer to Figure 1.18 on p 93.

Schelkunoff, S. A. and H.T. Friis, *Antennas: Theory and Practice*, Wiley, 1952.

Brown, G. H. and D.M. Woodward, "Experimentally Determined Radiation Characteristics of Conical and Triangular Antennas," *RCA Review*, Vol 13, no. 4, pp 425-452, Dec. 1952.

Nail, J. J., "Designing Discone Antennas," *Electronics*, Vol 26, Aug 1953, pp 167-169.

Crowley, T. H. and W. Marsh, *Discone Type Antennas*, technical report 510-13, Ohio State Research Foundation, 30 June 1954.

Wait, J. R., "Electromagnetic Radiation from Conical Structures," Chapter 12 in *Antenna Theory, Part 1*, R. E. Collin and F. J. Zucker (eds.), McGraw-Hill, 1969, ISBN 0070117993.

Lakshminarayana, Y., Y. R. Kubba, and M. E. Madhusudan, "A Wide Band Discone Antenna," *Indian Journal of. Electro-Technology*, Vol 15, no. 2, Mar/Apr 1971, pp 57-59.

Kadakia, D. R. and J. G. Fikioris, "Monopole Antenna Above a Hemisphere," *IEEE Transactions on Antennas and Propagation*, Vol 19, no. 5, Sep 1971, pp 687-690.

Spanos, W. M., *Broadband Discone Antenna having Auxiliary Cone*, US patent no. 3,618,107, 2 Nov 1971.

Hahn, R. F. and J. G. Fikioris, "Impedance and Radiation Pattern of Antennas above Flat Discs," *IEEE Transactions on Antennas and Propagation*, Vol 21, no. 1, Jan 1973, pp 97-100.

Woodman, K. F., "Dielectric-Clad Discone," *Electronics Letters*, Vol 13, no. 9, 28 Apr 1977, pp 264-265.

Smith, C. E., C. M. Butler, and K. R. Umashankar, "Characteristics of a Wire Biconical Antenna," *Microwave Journal*, Vol 22, no. 9, Sep 1979, pp 37-40.

Stutzman, W. L. and G. A. Thiele, *Antenna Theory and Design*, Section 6.2, Wiley, 1981, ISBN 047104458X.

McNamara, D. A., D. E. Baker and L. Botha,

"Some Design Considerations for Biconical Antennas," *IEEE AP-S Int'l Symp. Digest*, Vol 22, Jun 1984, pp 173-176.

Rappaport, T. S., *Tunable Discone Antenna*, US patent no. 4,851,859, 25 Jul 1989.

Austin, B. A. and A. P. C. Fourie, "Characteristics of the Wire Biconical Antenna Used for EMC Measurements," *IEEE Transactions on Electromagnetic Compatibility,* Vol 33, no. 3, Aug 1991, pp 179-187.

Champion, J. R., et al, *Portable Rapidly Erectable Discone Antenna*, US patent no. 5,608,416, 4 Mar 1997.

Samaddar, S. N. and E. L. Mokole, "Biconical Antennas with Unequal Cone Angles," *IEEE Transactions on Antennas and Propagation*, Vol 46, no. 2, Feb 1998, pp 181-193.

Meitzler, W., W. Turney, and L. T. Malek, "Biconical Antenna Scans 4.8 to 5.8 GHz," *Microwaves & RF*, Vol 44, no. 8, Aug 2005, pp 53-61.

Apostolos, J. T., *Collapsible Wide Band Width Discone Antenna*, US patent no. 6,967,626, 22 Nov 2005.

Marsan, L. A. and E. A. Urbanik, *Compact Low RCS Ultra-Wide Bandwidth Conical Monopole Antenna*, US patent no. 7,006,047, 28 Feb 2006.

Amateur Radio literature:

Boyer, J. M., W6UYH, "Discone — 40 to 500 Mc Skywire," *CQ*, Jul 1949, pp 11-15 & 69-71. Reprinted in *CQ Anthology*, Cowan Publishing, 1958.

Seybold, M., W2RYI, "The Low-Frequency Discone," *CQ*, July 1950, pp 13-16, 60. Reprinted in *CQ Anthology*, Cowan Publishing, 1958.

Orr, W. I., W6SAI, (ed.), *Radio Handbook*, 14th ed, "The Low-Frequency Discone," Editors and Engineers, 1956, pp 369-370 in.

Pappenfus, E. W., WB6LOH, "The Conical Monopole Antenna: Four-to-One Frequency Coverage with a Vertical," *QST*, Nov 1966, pp 21-24.

Wintzer, M., PA0MWI/DJ4GA, "Dipole Passe? Some Experiments on Discone Arrangements," *QST*, Oct 1974, pp 15-21.

Belrose, J. S., VE2CV/VE3DRC, "The HF Discone Antenna," *QST*, Jul 1975, pp 11-14, 56; Feedback, *QST*, Sep 1975, p 88.

Geiser, D., WA2ANU, "An Inexpensive Multiband VHF Antenna," *QST*, Dec 1978, pp 28-29.

Geiser, D., WA2ANU, "More on Discone Antennas," *QST*, May 1979, p 44.

White, T. E., K3WBH, "A Discone Antenna for 10 and 6 Meters and Lo-Band Public Service Monitoring," *CQ*, Jun 1980, pp 74-75.

Orr, W. I., W6SAI, "A 160-Meter Discone Antenna," *Ham Radio*, Jul 1987, pp 57-58.

Miller, D., NZ9E, "A 'Universal' VHF/UHF Antenna," *QST*, Dec 1994, p 75.

Krupp, D. A., W8NWF, "Two HF Discone Antennas," *The ARRL Antenna Compendium Vol* 5, ARRL, 1996, ISBN 0872595625, pp 121-127.

Todd, M., VK6JMA, "The Discone," *The International Antenna Collection*, G. Brown (ed.), RSGB and ARRL, 2003, ISBN 1872309933, pp 218-219.

Patterson, R., K5DZE, "A VHF/UHF Discone Antenna," *QST*, May 2003, pp 40-41. ⬛EX⊢

A Large Aperture, Resonant, Regenerative Frame Antenna (LARRFA)

A unique solution for HF reception in an antenna-restricted environment.

Bill Young, WD5HOH

Since December 2005 I've had a preliminary version of a large aperture, resonant, regenerative frame antenna "bench running" at home. It is capable of stable tuning and the level of feedback can easily be adjusted for maximum sensitivity just short of oscillation. The tuning range is somewhat limited (about 800 kHz — from about 5 MHz to about 5.8 MHz), but the level of performance is encouraging.

This project was, and continues to be, the result of necessity. I am not permitted to have any kind of external radio antenna. I have a short indoor whip antenna with a small home brew JFET preamp (thanks to *Popular Electronics*) which works well these days on 49 meters in the evenings, but I thought a tuned, regenerative frame antenna or loop might perform better.

A few years ago I built and operated a cascade regenerative receiver (*QEX* January/February, 2004) which has performed well. I knew then that I could operate a regenerative antenna ahead of a regenerative receiver. That was my starting point. I tried many circuits, most of them suggested by my newly acquired toy, *LTSPICE*. It took me several months to realize that (A) there are probably some things about the correct use of *LTSPICE* that have escaped my attention, and (B) just because a circuit appears to do just what you want it to do in *LTSPICE* does not mean it will actually perform as intended when built. The exercise was fun, but not productive.

4907 Grand Chateau Ln
Houston, TX 77084
bilaquieta@yahoo.com

Experiments Yield Results

So I then tried the most basic of circuits (Figure 1). I determined from my previous experience that the optimum inductive load for a regenerating 2N3819 is about 6500 Ω. I then placed a winding having that inductive reactance in the center of the intended frequency range just inside the three-turn tuned-gate coil. That's similar to what Edwin Howard Armstrong did when he first discovered regeneration. My circuit oscillated because I just happened to have the phase relationship of the two windings correct.

I have used the rectangular winding calculator at **emcsun.ece.umr.edu/new-induct/square.html** courtesy of the University of Missouri at Rolla to calculate frame winding inductance. The results have been good.

The next major task was to establish reliable, repeatable control of regeneration. Putting a potentiometer across the drain inductive load was not tried based on experience. See my *QEX* article, "A Mathematical Model for Regenerative RF Amplifiers" July/August,
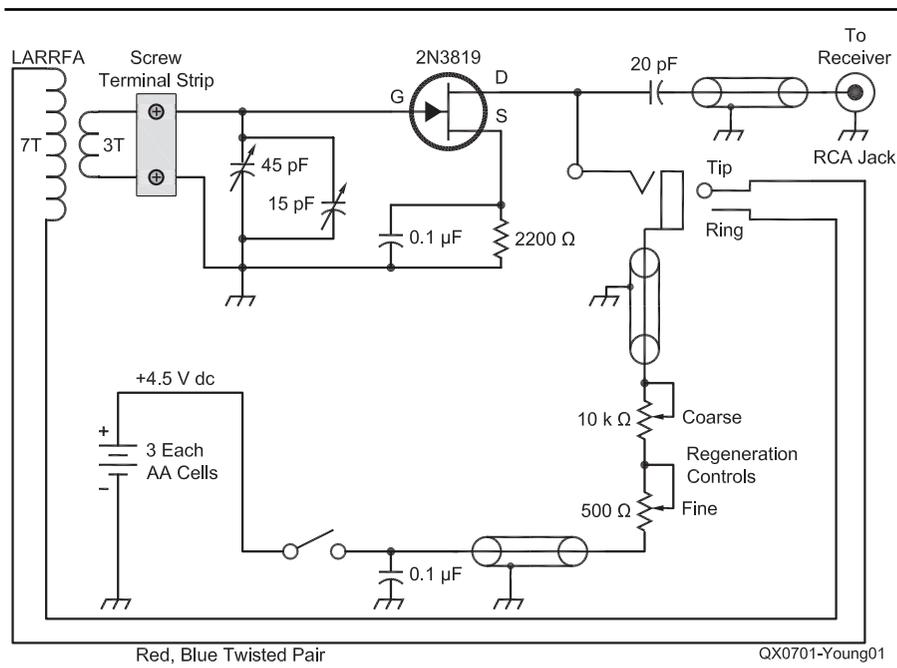
Figure 1—Schematic diagram of the LARRFA circuit.

Figure 2—The LARRFA frame with winding.

Labels in figure:
- 21 Inches
- 21 Inches
- Composition Board Stiffener, 2 each
- Turns spacing about 0.3 Inch
- 3 Turn signal winding
- 7 Turn drain load, tickler winding
- 0.75 Inch × 2.5 Inch Wood
- Side View
- Edge View
- NOTE: Leads not shown
- Spacing between the two windings approximately 0.5 Inch
- QX0701-Young02

The LARRFA Antenna

The frame antenna is three turns of AWG number 14 electrical wire wound on a wooden X frame approximately 20 inches on each side (see Figure 2). The three-turn gate winding is connected to the circuit chassis by a 10-inch length of 300 Ω twin lead. The tickler or feedback winding is a 7-turn winding of AWG 20 stranded geophysical wire (that I happened to have) wound in slots cut into the

2001. In that article I presented a mathematical model for a regenerative JFET circuit to demonstrate that changing the drain load impedance changes the rate of regeneration change. I also tried a 1 MΩ pot from drain to gate. Control was not repeatable and stable. The next configuration I tried is what I am continuing to use today: a 10 kΩ pot in series with a 500 Ω pot at the bottom "dc power supply end" of the JFET drain circuit. I see now that a 5 kΩ or even a 2.5 kΩ pot might be used for even smoother control. Supply voltage for the LARRFA circuitry is 4.5 V dc. Control of regeneration at this level is better than at 9 V.

You may wonder why I couldn't simply use one of the circuits for a regenerative receiver. The short answer appears to be that those circuits won't oscillate when the tuned inductance is wound on a wooden frame approximately 20 inches on a side. There are two reasons I can think of for this: (A) the Q of the inductance is lower than it was for the small solenoids and toroidal coils I have used for receivers, and (B) there may just be a small, but finite radiation resistance, which indicates that some energy is being radiated by the resonant coil (I'm not sure about this). Either or both of these effects would retard regeneration.

antenna frame just inside (toward the center) of the wooden X frame. It was necessary to tie these turns to the frame with lacing cord and hot glue to hold them in place.

One thing I need to do is position the antenna farther from the chassis. When I have the antenna and associated electronics peaked on a signal I notice that a change in my body position alters tuning and sensitivity. I may be able to accomplish that with four equal lengths of good quality coax.

The LARRFA electronics unit is built on a 7 × 5 × 3 inch aluminum chassis with the tuning capacitors and regeneration control potentiometers mounted on one side, which serves as a front panel. It's important to use coax for the regeneration control wiring inside the chassis as well as the signal output wiring inside the chassis. Failure to do this can result in erratic regeneration control. I ground one end only of each run of coax.

To operate the LARRFA, first tune the receiver to a strong signal. Switch the LARRFA system on with both regeneration controls fully clockwise (maximum regeneration). Turn the coarse regeneration control slowly counterclockwise in small increments, retuning the LARRFA main tuning control after each increment until a strong carrier whistle is heard in the connected receiver. It should be possible to continue this process until the LARRFA does not oscillate, but gain is produced. Retune the LARRFA until the signal of interest is "peaked." The very last regeneration adjustment should be made by turning the fine regeneration control counterclockwise.

Conclusion

The LARRFA is still very much a work in progress. I feel this antenna system may be of real interest to people who want to receive, but who can't have an outdoor antenna. Also, I think it's very likely the LARRFA can be used with just about any communications receiver, regenerative or otherwise. A well-designed source follower circuit would help by matching the LARRFA output impedance more closely to the typical 75 to 50 Ω communications receiver input.

I have been encouraged by an article by Dan Wissell describing a high frequency loop antenna (not regenerative). I have taken this idea one step further and I am not alone. There are other regenerative loop antennas described on the Web.

The most important unresolved issue with respect to the LARRFA is this: Does it capture enough signal to be of real use as a short wave antenna? I know it has high gain based on the consistently high man-made background noise levels I hear when it's peaked to the receiver's tuned frequency. And I know it does capture some signals such as WWV and various digital signals. Further operation will tell. **QEX**

On the Crossed Field Antenna Performance, Part 1

Valentino Trainotti, *Senior Member, IEEE*, **and**
Luis A. Dorado, *Student Member, IEEE*

Abstract— Lately, short antennas and Crossed Field Antennas (CFA) have attracted broadcast and amateur community attention.

The CFA antenna has been developed in the last decade of the 20th century, trying to obtain a compact transmitting antenna for low and medium frequency AM bands. The CFA is intended to be used in order to get a low profile antenna and a supposed performance similar or better compared to a quarter-wave monopole.

The CFA has a short monopole and a metallic disk close to the monopole base, both mechanical structures being fed by means of two separated generators. Thus, the CFA has two ports and can be analyzed from the Network Theory point of view.

In this paper, the CFA has been studied exhaustively using the Transmission Line Method (TLM) in order to obtain an equivalent network and the antenna performance. Due to the lack of theoretical data to explain the CFA antenna behavior, the TLM has been validated by means of Moment Method simulations and some available experimental data.

As a first approximation, the CFA is placed on a perfectly electric conducting (PEC) ground plane in order to obtain the antenna currents and near fields. Once this task has been performed, losses due to an actual ground and an artificial metallic ground plane, that is of common use in practice, can be calculated.

The novel approach here permits to obtain the near and far field expressions from the current distributions on the antenna structure. Then, near field calculations are used to determine the surface current density on the ground plane around the antenna and the wave impedance as a function of distance in space.

Near fields and wave impedance are used for determining whether the Poynting Vector Synthesis (PVS) phenomenon exists or not. PVS means that the far field zone boundary is located at the surface of the CFA antenna structure itself, the power density or Poynting vector being real everywhere in space.

From the artificial and natural ground plane surface current density, the power dissipation is calculated in a circular boundary half-wavelength from the antenna base, and the ground plane equivalent loss resistance is obtained.

Artificial ground plane behavior is paramount in obtaining the best performance of a CFA antenna due to its short height in wavelengths, as well as in any monopole antenna type.

Index Terms— CFA, Crossed Field Antenna, CFA Antenna, LF Antennas, LF Broadcast Antennas, LF Broadcast Transmitting Antennas, LF AM Broadcast Antennas, LF Monopole Antennas, MF Antennas, MF Broadcast Antennas, MF Broadcast Transmitting Antennas, LF Short Transmitting Antennas, MF Short Transmitting Antennas, MF AM Broadcast Antennas, MF Monopole Antennas, Short monopole, Top-loaded monopole, Ground plane, Grounding, Antenna Input Impedance, Antenna Efficiency, Antenna Gain, Antenna Performance, Antenna Bandwidth.

University of Buenos Aires,
Argentina
vtrainotti@ieee.org
luis_dorado@ieee.org
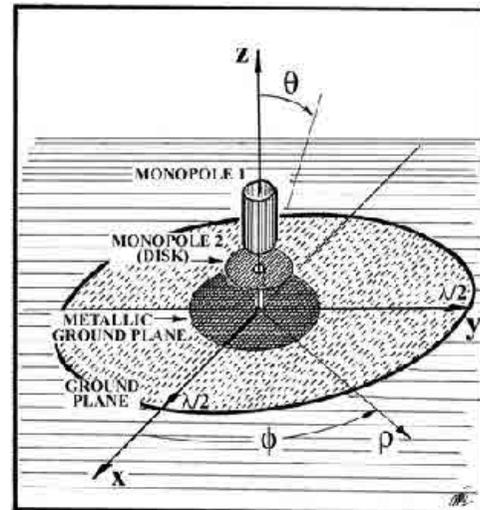
Fig. 1.   Crossed Field Antenna (CFA) general sketch.

I. Introduction

ANY linear antenna is a Crossed Field Antenna (CFA) in the far field zone, in the sense that the electric field is perpendicular to the magnetic field in space and both fields are perpendicular to the direction of wave propagation, which is indicated by a real power density or Poynting vector.

Nevertheless, the term CFA was intended to be used for an antenna where the existence of the crossed fields is everywhere in space, which is known as Poynting Vector Synthesis (PVS). Calculations demonstrate that this is a utopia, because on the antenna metallic structure both fields must fulfill the boundary conditions and a near field zone always exists, where induction fields predominate.

The CFA antenna has a short monopole (monopole 1) and another very short monopole (monopole 2) with a metallic disk as a top-load, which is parallel and close to the earth. The dimensions of both metallic structures, monopoles 1 and 2, are much smaller than the wavelength, then the Transmission Line Method (TLM) for the analysis of top-loaded monopoles applies [1]. The monopole 1 axis passes through a hole at the disk center, as can be seen in Fig. 1.

Generally, the ground below the antenna structure is covered by a thin metallic layer whose diameter is greater than the metallic disk.

Input power is injected into the antenna by means of two generators, the first one is connected to the monopole 1 base and ground and the second one is connected to the monopole 2 (disk) and ground.

In this paper, the CFA antenna was thoroughly analyzed from the input impedance and radiation properties points of view using Maxwell equations through the TLM and a
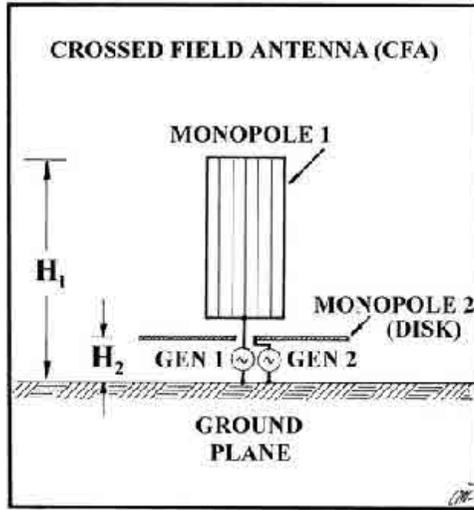
Fig. 2. Crossed Field Antenna (CFA) feeding system.
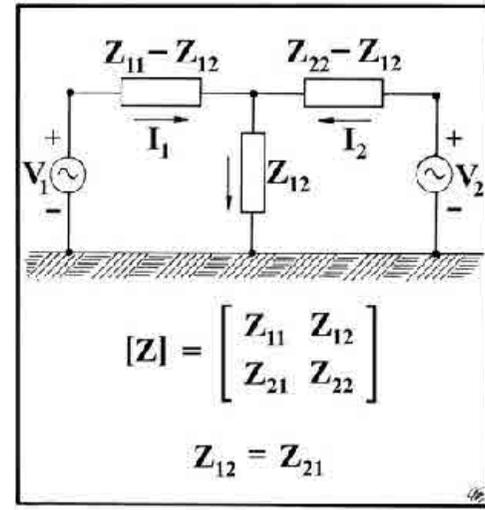
Fig. 3. Crossed Field Antenna (CFA) equivalent network.

Moment Method approach. No serious theoretical analysis of this antenna has been found in the technical literature and only partial analysis made up by software simulations [2]–[5] or some experiment with actual and reduced scale models [6]–[10]. Both intents do not explain clearly the actual antenna performance, inventors explanations are too vague and they do not support their invention by means of a clear theory in order to assure its working performance [11]–[16].

II. EQUIVALENT NETWORK

The CFA antenna can be analyzed from a network point of view, taking into account the two ports in the antenna structure. Monopole 1 base and ground are the port 1 terminals, while metallic disk (monopole 2) and ground are the port 2 terminals. A sketch of this antenna can be seen in Fig. 2.

Generators are connected to both monopoles, 1 and 2 (disk), with short leads as it is usual in high and very high frequency techniques. Pictures of some published CFA antennas have their feeding connections with long wires placed in the space occupied by the near fields. It seems that this technique is the same as used by electrical 50 or 60 Hz installations and it should be avoided in RF frequency systems.

The antenna equivalent network can be characterized by its impedance parameters, the monopole 1 self-impedance, Z_{11}, the monopole 2 self-impedance (disk), Z_{22}, and the mutual impedances, Z_{12} and Z_{21}, between them.

Fig. 3 shows the antenna equivalent network, where a generator is connected to each port. Generator 1, connected to port 1, has a voltage V_1, while generator 2, connected to port 2, has a voltage V_2. As soon as both generators are connected to the antenna ports, currents I_1 and I_2 will flow into the antenna structure.

By using impedance parameters, the standard network equations can be written in matrix form as follows [17]–[19]:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad (1)$$

From the Reciprocity Theorem, it follows that $Z_{12} = Z_{21}$.

Admittance parameters will also be required and are obtained by inverting the impedance matrix, that is

$$Y_{11} = G_{11} + B_{11} = \frac{Z_{22}}{Z_{11} Z_{22} - Z_{12}^2} \quad (2)$$

$$Y_{12} = Y_{21} = G_{12} + B_{12} = -\frac{Z_{12}}{Z_{11} Z_{22} - Z_{12}^2} \quad (3)$$

$$Y_{22} = G_{22} + B_{22} = \frac{Z_{11}}{Z_{11} Z_{22} - Z_{12}^2} \quad (4)$$

The impedance and admittance matrices are functions of frequency and the physical dimensions of the antenna system.

A. Network Parameters

The CFA antenna can be considered as an array of two tightly coupled short monopoles. The first one is a monopole of height H_1, which could have or not a top-load, while the second one is a monopole of height H_2 with the disk as its top-load. These monopoles are shown in Figs. 4 and 5.

Thus, the present theory is based on the Transmission Line Method (TLM) outlined in [1] for the analysis of a top-loaded monopole [20].

When port 2 is an open circuit, the monopole 1 input impedance, $Z_{11} = R_{11} + X_{11}$, is the impedance of a short monopole with a given degree of top-loading.

When port 1 is an open circuit, the monopole 2 input impedance, $Z_{22} = R_{22} + X_{22}$, is the impedance of another very short monopole with the disk as its top-load (practically a Hertz monopole).

Self-resistance and reactance of each monopole are given by [1]

$$R_{ii} = R_{radi} + R_{ci} + R_{gpi} \qquad = 1 \; 2 \quad (5)$$

$$X_{ii} = Z_{0mi} \frac{Z_{0mi} \quad H_i + X_{ti}}{Z_{0mi} - X_{ti} \quad H_i} \qquad = 1 \; 2 \quad (6)$$
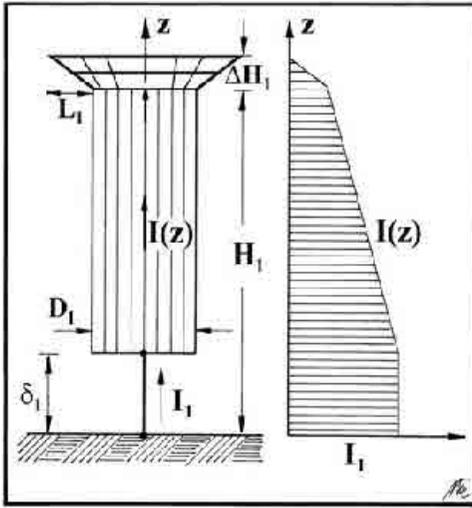
Fig. 4. CFA monopole 1 current distribution.
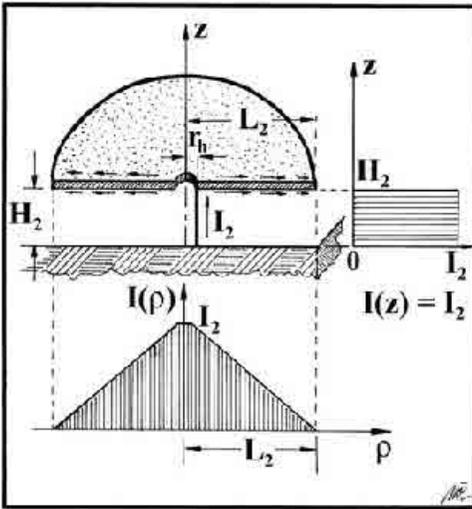
Fig. 5. CFA monopole 2 and top-loading disk current distributions.

Where

R_{radi} is the i-th monopole radiation resistance .

R_{ci} is the i-th monopole conductor resistance .

R_{gpi} is the i-th monopole ground plane loss resistance .

Z_{0mi} is the average characteristic impedance of the ith-monopole equivalent transmission line [1] .

X_{ti} is the ith-monopole top-reactance .

H_i is the ith-monopole height [m].

The i-th monopole radiation resistance is [21]–[23]

$$R_{radi} = 40\left(\ H_{ei}\right)^2 \qquad = 1\ 2 \qquad (7)$$

Where H_{ei} is the i-th monopole effective height, which depends on the vertical wire current distribution, and it is given by

$$H_{ei} = \quad H_i + \frac{X_{ii}}{Z_{0mi}}\left(1 - \quad H_i\right) \qquad = 1\ 2 \quad (8)$$

Due to a very small disk height, H_2, over the ground plane and a disk radius, L_2, greater than this height, the current distribution on monopole 2 is practically constant, therefore

$$H_{e2} \cong H_2 \qquad (9)$$

The top-reactance of monopole 1, X_{t1}, is [1]

$$X_{t1} = -\frac{Z_{0t1}}{\quad L_1} \qquad (10)$$

Where

Z_{0t1} is the monopole 1 top-load characteristic impedance [1] .

L_1 is the monopole 1 top-load length [m].

is the number of top-load branches of monopole 1.

The top-reactance of monopole 2, X_{t2}, is

$$X_{t2} = -\frac{1}{2\quad C_2} \qquad (11)$$

Where the disk capacitance, C_2, can practically be calculated as a simple capacitor with a circular plate of radius L_2 and a separation from the ground plane equal to the monopole 2 height, H_2, then

$$C_2 = \ _0\frac{L_2^2}{H_2} \qquad (12)$$

When port 2 is an open circuit, the mutual impedance Z_{21} is the ratio between the open circuit voltage in port 2 and the current flowing in port 1, that is

$$Z_{21} = \left.\frac{V_2}{I_1}\right|_{I_2 = 0} \qquad (13)$$

Also, the open circuit voltage V_2 is due to the electric field E_{z1} between the capacitor plate and ground [24], according to the current I_1 in port 1, then

$$Z_{21} = -\left.\frac{E_{z1}\,H_2}{I_1}\right|_{I_2 = 0} \qquad (14)$$

The electric field E_{z1} is the near field produced by monopole 1 in the space surrounding the antenna and can be calculated by means of (22).

Another possibility to calculate the impedance and admittance parameters is using the Method of Moments (MoM), which has been used extensively by means of our own software [25], but similar results could be obtained by means of other standard softwares. These parameters are practically the same for both techniques, Transmission Line Method (TLM) and MoM, as practical examples indicate.

Nevertheless, it is important to have at hands the electromagnetic equations in order to solve the problem and have a clear view of the antenna behavior.

B. Input Impedances

The antenna is excited by means of two generators, for this reason, monopole 1 input voltage V_1 will be taken as the phase reference and monopole 2 input voltage V_2 will be given by

$$V_2 = K V_1^{\ j\phi_2} \qquad (15)$$

Thus, the parameter K is the amplitude ratio between V_1 and V_2, while ϕ_2 is the phase difference between them. Voltages and currents are taken as effective values.

The port 1 input impedance is

$$Z_1 = R_1 + X_1 = \frac{V_1}{I_1} = \frac{1}{Y_{11} + Y_{12} K^{\ j\phi_2}} \qquad (16)$$

The port 2 input impedance is

$$Z_2 = R_2 + X_2 = \frac{V_2}{I_2} = \frac{K}{Y_{12}^{\ -j\phi_2} + K Y_{22}} \qquad (17)$$

It can be seen that the input impedances depend on a strong interaction between both generators.

C. Input Power

The active or real power $W_1 = |I_1|^2 R_1$, produced by generator 1, is given by

$$W_1 = |V_1|^2 \ (G_{11} + K G_{12} \ \ \phi_2 - K B_{12} \ \ \phi_2) \qquad (18)$$

The active or real power $W_2 = |I_2|^2 R_2$, produced by generator 2, is given by

$$W_2 = |V_1|^2 \ (K^2 G_{22} + K G_{12} \ \ \phi_2 + K B_{12} \ \ \phi_2) \qquad (19)$$

Then, the total input power will be the sum of the input powers of both generators, $W_{in} = W_1 + W_2$, that is

$$W_{in} = |V_1|^2 \ (G_{11} + 2 K G_{12} \ \ \phi_2 + K^2 G_{22}) \qquad (20)$$

Equation (20) shows that the input power depends on the cosine of the voltage phase difference ϕ_2. For this reason, there are three cases:

(I) When $G_{12} \quad 0$, the input power is maximum for $\phi_2 = 0 \ (360°)$ and minimum for $\phi_2 = 180°$.

(II) When $G_{12} \quad 0$, the input power is maximum for $\phi_2 = 180°$ and minimum for $\phi_2 = 0 \ (360°)$.

(III) When $G_{12} = 0$, the input power is constant for any value of ϕ_2.

These three cases or regimes depend on the antenna physical dimensions and frequency, and once these are given, the antenna will operate in only one of those regimes.

In Fig. 6, a CFA operating in the first regime is shown ($G_{12} \quad 0$, $\phi_2 = 0$) as well as the antenna equivalent circuit, where the currents of both generators are in phase. In Fig. 7 a CFA operating in the second regime is shown ($G_{12} \quad 0$, $\phi_2 = 180°$), where the generator currents are out of phase. When $G_{12} = 0$, the CFA will operate either as in Figs. 6 or 7 according to the highest voltage generator, however, the total input power will be independent of ϕ_2.
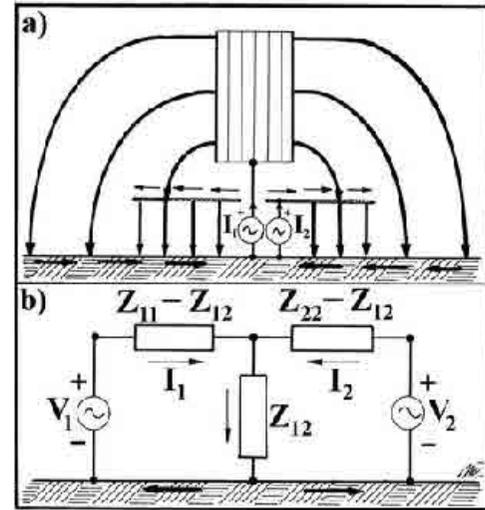
Fig. 6.   a) CFA operating in the first regime (I) $G_{12} > 0$. b) CFA equivalent circuit.
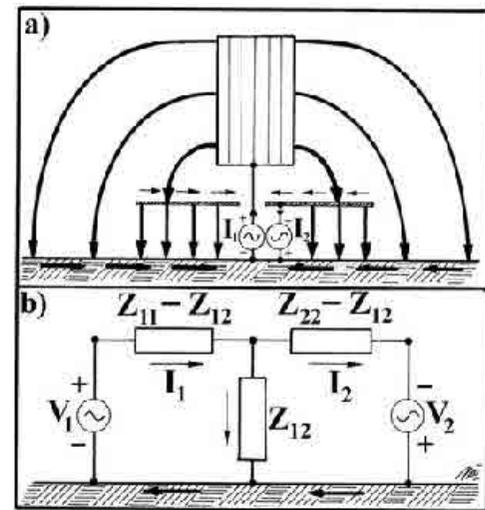
Fig. 7.   a) CFA operating in the second regime (II) $G_{12} < 0$. b) CFA equivalent circuit.

III. ELECTROMAGNETIC FIELD

The electromagnetic field radiated from the antenna is the sum of the radiation produced by the array of the two CFA monopoles. Monopole 1 is a top-loaded short vertical antenna and monopole 2 is practically a Hertz monopole makes up by the disk and its feeding vertical lead.

Because of the very short distance between both radiators, the radiation center is located on the CFA vertical geometric axis at a zero height. The feeding currents of each radiating source have different amplitudes and phases depending on the voltages of both generators.

A. Near Field

The near field of a top-loaded monopole has been determined in [1] using Image Theory, according to Fig. 8 geometry.

Then, the magnetic and electric near fields of both monopoles on the ground plane, at $= 0$, are given by [1]
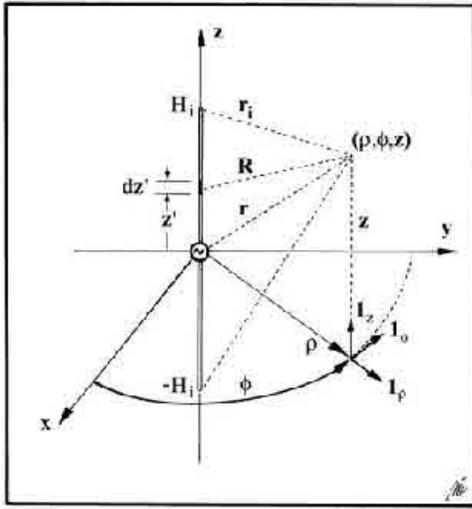
Fig. 8. Monopole geometry used to calculate the near field in cylindrical coordinates.

$$H_{\phi i} = -\frac{I_{mi}}{4\pi\rho}\left\{e^{j\psi_i}\left[\left(1-\frac{H_i}{r_i}\right)e^{-j\beta(r_i+H_i)} - e^{-j\beta\rho}\right]\right.$$
$$\left. - e^{-j\psi_i}\left[\left(1+\frac{H_i}{r_i}\right)e^{-j\beta(r_i-H_i)} - e^{-j\beta\rho}\right]\right\} \quad (21)$$

$$E_{zi} = \frac{I_{mi}}{4\pi\varepsilon_0\omega\rho} \quad (22)$$
$$\left\{e^{j\psi_i}\left[\frac{\rho\,e^{-j\beta(r_i+H_i)}}{r_i}\left(\frac{H_i}{r_i^2} - j\beta\left(1-\frac{H_i}{r_i}\right)\right)\right.\right.$$
$$\left. + j\beta\,e^{-j\beta\rho}\right] +$$
$$+ e^{-j\psi_i}\left[\frac{\rho\,e^{-j\beta(r_i-H_i)}}{r_i}\left(\frac{H_i}{r_i^2} + j\beta\left(1+\frac{H_i}{r_i}\right)\right)\right.$$
$$\left.\left. - j\beta\,e^{-j\beta\rho}\right]\right\}$$

Where $i = 1, 2$ and

$$r_i = \sqrt{\rho^2 + H_i^2}$$

$$I_{mi} = I_i\sqrt{1+\left(\frac{X_{ii}}{Z_{0mi}}\right)^2}$$

$$\psi_i = \text{arctan}\left(\frac{X_{ii}}{Z_{0mi}}\right)$$

It is assumed, as in [1], that only the vertical part of a top-loaded monopole produces a net electromagnetic field. The distance ρ along the surface of the earth is measured from the center of the cylindrical coordinates or antenna vertical axis.

The CFA total near electromagnetic fields on the ground plane ($z = 0$) will be

$$H_\phi = H_{\phi 1} + H_{\phi 2} \quad (23)$$

$$E_z = E_{z1} + E_{z2} \quad (24)$$

On the surface of the earth, a thin metallic layer of radius R_0 is laid down, which is called the *artificial ground plane*. From this metallic layer radius R_0, the bare soil is considered up to a distance of half-wavelength. This is the surface of the earth to be taken into account in the soil power loss calculations.

Over the metallic layer or artificial ground plane the electric field is practically perpendicular, having only the E_z component, due to the boundary conditions of a very high conductivity medium, where the surface impedance Z_g is given by [26], [27]

$$Z_g = R_g + jX_g = \sqrt{\frac{\omega\mu_0}{2\sigma_m}}(1+j) \quad \text{for} \quad \rho \leq R_0 \quad (25)$$

Where σ_m is the metallic layer conductivity.

On the bare soil, the electric field develops a radial component E_ρ due to the low conductivity compared to that of the metallic layer, and is given by

$$E_\rho = -Z_s H_\phi \quad (26)$$

Where Z_s is the bare soil impedance, which can be calculated from the soil conductivity, σ, and permittivity, ε, as follows [26], [27]

$$Z_s = R_s + jX_s = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\varepsilon}} \quad \text{for} \quad \rho \geq R_0 \quad (27)$$

B. Wave Impedance

The wave impedance in space, just above the earth surface in the air, is the ratio between the near electric and magnetic fields E_z and H_ϕ on the ground plane, and is a function of distance ρ, therefore

$$Z_0 = -\frac{E_z}{H_\phi} \quad (28)$$

In the far field zone, $\rho \gg \lambda$, the wave impedance is practically equal to the free space intrinsic impedance $Z_{00} = 120\pi$. The wave impedance Z_0 is almost pure imaginary or reactive very close to the antenna and almost pure real or resistive at a distance greater than half-wavelength.

C. Far Field

In the far field zone, when spherical coordinates are used, the CFA total far electric field E_θ will be the sum of the monopoles 1 and 2 far electric fields, $E_{\theta 1}$ and $E_{\theta 2}$, which are given by [22], [23]

$$E_{\theta i} = j60\,I_i\,H_{ei}\frac{e^{-j\beta r}}{r} \quad i = 1, 2 \quad (29)$$

Where

r is the distance from the antenna to any point in space [m].
is the zenith angle.

The total far electric field, $E_\theta = E_{\theta 1} + E_{\theta 2}$, will be

$$E_\theta = 60 \left(I_1 \, H_{e1} + I_2 \, H_{e2} \right) \frac{e^{-j\beta r}}{1} \qquad (30)$$

Then

$$| E_\theta | = 60 | I_1 | \, H_e \frac{1}{} \qquad (31)$$

Where the CFA effective height H_e is defined as

$$H_e = H_{e1} \left| 1 + \frac{I_2 \, H_{e2}}{I_1 \, H_{e1}} \right| \qquad (32)$$

Therefore, the CFA radiation resistance referred to monopole 1 can be defined as

$$R_{rad} = 40 \left(H_e \right)^2 \qquad (33)$$

or

$$R_{rad} = 40 \left(H_{e1} \right)^2 \left| 1 + \frac{I_2 \, H_{e2}}{I_1 \, H_{e1}} \right|^2 \qquad (34)$$

It is very important to understand that this radiation resistance is not the input resistance in any of the antenna ports and it depends on the current and effective height relationships of both monopoles.

The antenna radiated power can be calculated as

$$W_{rad} = | I_1 |^2 \, R_{rad} \qquad (35)$$

Then, monopole 1 effective input current $|I_1|$, for a given radiated power W_{rad}, becomes

$$| I_1 | = \sqrt{\frac{W_{rad}}{R_{rad}}} \qquad (36)$$

Equations (31), (33) and (36) give the following far effective electric field:

$$| E_\theta | = \frac{\sqrt{30 \, W_{rad} \, D}}{1} \qquad (37)$$

Thus, the CFA antenna directivity is $D = 3$ due to the 1 far field radiation pattern, as given by any short monopole of height less than 0 1 , as was confirmed by antenna pattern measurements using a model [6].

Taking into account the antenna efficiency , , the radiated power W_{rad} and gain G are given by [26]

$$W_{rad} = W_{in} \qquad (38)$$

$$G = D \qquad (39)$$

Then, the CFA far effective electric field becomes

$$| E_\theta | = \frac{\sqrt{30 \, W_{in} \, G}}{1} \qquad (40)$$

This equation is exactly the same as for any standard short monopole. This field is the non attenuated radiated electric field, because it depends only on the inverse distance law. The actual field intensity along the earth is affected by the soil physical constants and the diffraction due to the spherical earth [28], [29].

IV. ANTENNA TUNING AND LOSSES

In the antenna circuit there are losses in conductors, insulators, tuning system and in the earth surface within a circle half-wavelength in radius. In general, insulator losses are very low compared to the other and can be neglected.

In order to tune the antenna, a coil is connected in series to each port, obtaining real input impedances. These coils have merit factors given by

$$Q_{Li} = \frac{X_{Li}}{R_{Li}} \qquad 1 = 1 \quad 2 \qquad (41)$$

Where R_{Li} and X_{Li} are the i-th monopole tuning coil resistance and reactance, respectively.

Network parameters can be written in the following way when losses and tuning coils are present:

$$Z_{ii} = Z_{ii}^\infty + R_{ci} + R_{gpi} + R_{Li} + X_{Li} \qquad 1 = 1 \quad 2 \quad (42)$$

$$Z_{12} = Z_{21} = R_{12} + X_{12} \qquad (43)$$

Where
$Z_{ii}^\infty = R_{radi} + X_{ii}$ is the i-th monopole self-impedance with no losses and no tuning coils [ ].

R_{radi} is the i-th monopole radiation resistance [ ].

R_{ci} is the i-th monopole conductor resistance [ ].

R_{gpi} is the ground plane loss resistance due to the i-th monopole current distribution and soil conditions [ ].

The mutual resistance R_{12} and all the reactances are practically not affected by the antenna losses, because they depend on the very near field distribution, which is not appreciably affected by the finite soil conductivity [30], [31].

The CFA equivalent network with losses and tuning coils can be seen in Fig. 9.

A. Conductor Loss Resistance

Conductor loss resistance R_{c1}, due to monopole 1 conductors, can be calculated using the following expression, determined from the current distribution on the monopole structure and its top-load [1]:
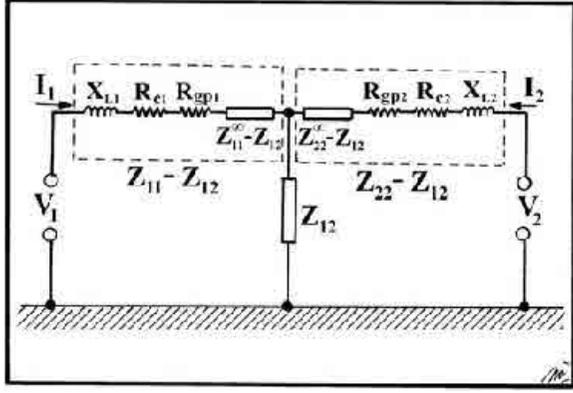
Fig. 9. CFA equivalent network with losses and tuning coils.
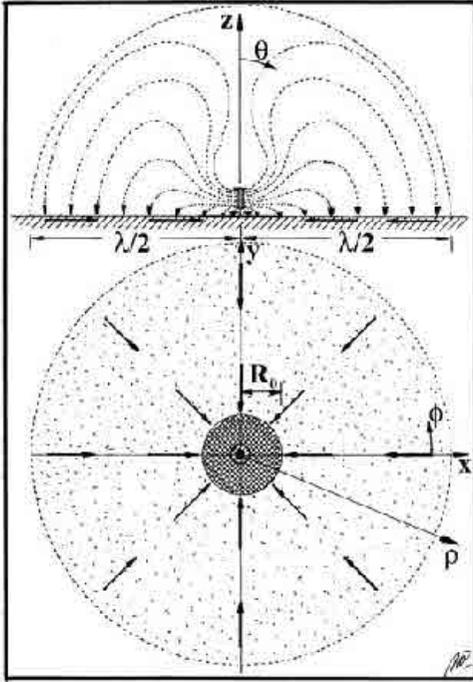
Fig. 10. CFA near electric field lines (displacement current) and ground plane surface conductive currents.

$$R_{c1} = \frac{R_{cL1}}{2}\left\{ \left(1 + \frac{X_{11}^2}{Z_{0m1}^2}\right) H_1 + \right.$$

$$\left(1 - \frac{X_{11}^2}{Z_{0m1}^2}\right)\frac{\sin 2\beta H_1}{2\beta} + \frac{X_{11}}{Z_{0m1}}\frac{1 - \cos 2\beta H_1}{\beta} +$$

$$\frac{1}{n}\left(\cos\beta H_1 + \frac{X_{11}}{Z_{0m1}}\sin\beta H_1\right)^2$$

$$\left[L_1\left(1 + \frac{1}{\tan^2\beta L_1}\right) + \right.$$

$$\left. \frac{\sin 2\beta L_1}{2\beta}\left(1 - \frac{1}{\tan^2\beta L_1}\right) + \frac{\cos 2\beta L_1 - 1}{\beta\tan\beta L_1}\right]\right\} \quad (44)$$

Where

R_{cL1} is the monopole 1 wire resistance per unit length [Ω/m].

X_{11} is the monopole 1 self-reactance [Ω].

Z_{0m1} is the monopole 1 average characteristic impedance [Ω].

H_1 is the monopole 1 height [m].

L_1 is the monopole 1 top-load length [m].

n is the number of the monopole 1 top-load branches.

The current distribution of the top-loaded monopole has been determined in [1] and can be seen in Fig. 4.

Conductor resistance per unit length, taking into account the skin effect, is given by [26]

$$R_{cL1} = \frac{1}{a_1}\sqrt{\frac{f\,\mu_0}{4\,\pi\,\sigma_c}} \quad (45)$$

Where

σ_c is the conductor conductivity [S/m].

a_1 is the wire radius [m].

f is the operation frequency [Hz].

If the disk radius L_2 is sufficiently smaller than the wavelength, a linear current distribution over the disk can be assumed, as shown in Fig. 5, therefore

$$I(\rho) = I_2\frac{L_2 - \rho}{L_2 - r_h} \qquad r_h \le \rho \le L_2 \quad (46)$$

Where

$I(\rho)$ is the current distribution over the disk [A].

I_2 is the effective current at the disk center [A].

L_2 is the disk radius [m].

r_h is the disk hole radius [m].

The disk feeding lead has a constant current distribution like a Hertz monopole ($I(z) = I_2$ in Fig. 5). From the disk and feeding lead current distributions, the monopole 2 conductor resistance, R_{c2}, can be calculated (see Appendix A), therefore

$$R_{c2} = R_{cL2}H_2 + \frac{1}{4\pi(L_2 - r_h)^2}\sqrt{\frac{\omega\,\mu_0}{2\,\sigma_c}}$$

$$\left[L_2^2\ln\left(\frac{L_2}{r_h}\right) - 2L_2(L_2 - r_h) + \frac{L_2^2 - r_h^2}{2}\right] \quad (47)$$

The feeding lead conductor resistance per unit length, R_{cL2}, is given by (45) with subscript 2 instead of 1.

B. Ground Plane Loss Resistance

The artificial ground plane [32] is made up by means of a circular metallic layer R_0 in radius with a small surface resistance R_g. Then, the ground plane equivalent loss resistance for each monopole is given by [1]

$$R_{gpi} = \frac{2\pi}{|I_i|^2}\left(\int_0^{R_0}|H_{\phi i}|^2 R_g\,\rho\,d\rho + \int_{R_0}^{\lambda/2}|H_{\phi i}|^2 R_s\,\rho\,d\rho\right) \quad (48)$$

Where i = 1, 2 and

R_{gpi} is the i-th monopole ground plane loss resistance [Ω].

$H_{\phi i}$ is the i-th monopole near magnetic field on the ground surface, given by (21) [A/m].

R_g is the surface resistance of the artificial ground plane or metallic layer, given by (25) [Ω].

R_s is the surface resistance of the natural ground plane or soil, given by (27) [Ω].

Fig. 10 shows a sketch of the displacement currents and the conductive currents flowing on the ground plane at a specific time.

The ground plane is considered up to a distance $\rho = \lambda/2$ because this is the maximum distance covered by the ground surface currents under the antenna, closing the antenna electric circuit. Beyond this distance, the ground currents do not return to the antenna generators and are taken into account in the surface wave propagation calculations [1].

V. EFFICIENCY AND GAIN

Antenna efficiency η is defined as the ratio between the radiated power W_{rad} and the input power W_{in}. The power radiated by the CFA antenna is given by (35), while the antenna input power can be written as

$$W_{in} = |I_1|^2 R_1 + |I_2|^2 R_2 \qquad (49)$$

Where R_i is the i-th port input resistance, i = 1, 2. Then, the CFA efficiency will be

$$\eta = \frac{R_{rad}}{R_1 + |I_2/I_1|^2 R_2} \qquad (50)$$

Where

$$\left| \frac{I_2}{I_1} \right|^2 = K^2 \left| \frac{Z_1}{Z_2} \right|^2 \qquad (51)$$

and the input impedances Z_1 and Z_2 are given by (16) and (17), respectively.

It can be seen that the CFA efficiency η depends on the antenna geometry, operation frequency, the voltage amplitude ratio K and the phase difference ϕ_2.

Antenna gain, taking into account the directivity D = 3, is $G = \eta D = 3\eta$.

VI. BANDWIDTH

The CFA antenna has two ports, for this reason, two bandwidths can be defined, according to the reflection coefficients of both ports. In this case, taking into account that both input powers, W_1 and W_2, must be positive, it is more convenient to define a CFA average reflection coefficient as a weighted geometric mean of the reflection coefficients, Γ_1 and Γ_2, of both ports. Thus, the CFA average reflection coefficient is given by

$$|\Gamma| = \sqrt{\frac{W_1 |\Gamma_1|^2 + W_2 |\Gamma_2|^2}{W_{in}}} \qquad (52)$$

Then, the CFA VSWR will be

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad (53)$$

Part 2 of this discourse will be presented in the next issue of QEX, *which will include measured parameters of the CFA.* QEX

Antenna Options

L. B. Cebik, W4RNL
1434 High Mesa Drive
Knoxville, TN 37938-4443
cebik@cebik.com

Narrowband NVIS Antennas

Dean Straw, N6BV, wrote an article for *QST* in December, 2005: "What's the Deal About 'NVIS'?" The article provides some excellent guidance for obtaining the best results from Near Vertical Incidence Skywave (NVIS) operation. The discussion limits itself to using a simple inverted-V antenna, which prompted the following notes. We have a number of options for potentially effective NVIS antennas. In this episode, we shall look at antennas that are narrow band, that is, antennas that cover one or part of one amateur band. We have enough to learn about them to occupy us fully.

Figure 1 sketches the NVIS situation in the most general terms. Regular amateur operations seek to elevate antennas to provide low-angle radiation. Ionospheric refraction results in a skip zone — an area between the central station and the nearest communications target. In addition, many central stations have obstructions that limit the range of point-to-point communications methods. In both cases, directing a lower-frequency HF signal upward can result in a sufficient return to provide short to intermediate range communications. Many government services consider the NVIS frequency range to extend from 2 to about 10 MHz. As Straw notes, the 7-MHz region is most suitable for nighttime work, while the 80/75-meter band provides the best results for daytime operation by radio amateurs.

Although you may set up many antennas for somewhat directional patterns, most operators strive to have an omni-directional antenna. Unfortunately, pure omni-directionality is hard to obtain with simple antennas. You can approximate a circular azimuth pattern, however, by choosing the right antenna, as shown on the left in Figure 2. The elongated azimuth pattern shown on the right may also be useful. For simple wire antennas, the broader pattern is off the ends of the wire. If the installation area permits, you can go some distance in planning your coverage. The figure also shows a convention that I shall use in these notes: listing the broadside and the endwise half-power beamwidth. The closer these numbers are to each other, the more circular will be the pattern. The greater the difference, the more elongated that the oval pattern becomes. The ratio of one to the other is a useful measure of pattern circularity.

The final preliminary note concerns the antenna environment. We shall be looking at narrow-band antennas for use at the central station. In amateur terms, that generally means a durable home installation for which one may plan and then construct with care. The short-masted AS-2259 antenna is designed for field use, which might be a central station on a military battlefield. The antenna is useful to amateurs in Field Day and similar exercises. For long-term NVIS antennas, however, we can do far better.

The Lowly Dipole

The standard AWG no. 14 copper wire dipole will be our starting point for these *NEC-4* modeling tests. All antennas will use average ground (conductivity = 0.005 S/m, relative permittivity = 13). The test frequency

Table 1
Modeled Performance of an NVIS Dipole Over Bare Ground at Various Heights

Height (λ)	NEC-4 Gain (dBi)	Beamwidth (Degrees)		MININEC Gain (dBi)
		Broadside	Endwise	
0.05	0.98	98	71	9.20
0.75	3.51	99	67	8.72
0.1	4.30	102	65	8.32
0.125	5.80	105	65	7.96
0.15	6.23	108	65	7.61
0.175	6.40	113	66	7.26
0.2	6.39	118	68	6.88
0.225	6.25	124	70	6.46
0.25	5.97	129	73	5.99

Notes
1. Antennas use AWG #14 copper wire over average ground (conductivity = 0.005 S/m, relative permittivity = 13).
2. *NEC-4* gain values use the Sommerfeld-Norton (S-N, called "High Accuracy" in *EZNEC*) ground calculation system. *MININEC* uses a simplified reflection coefficient approximation and applies it to far-field data only.
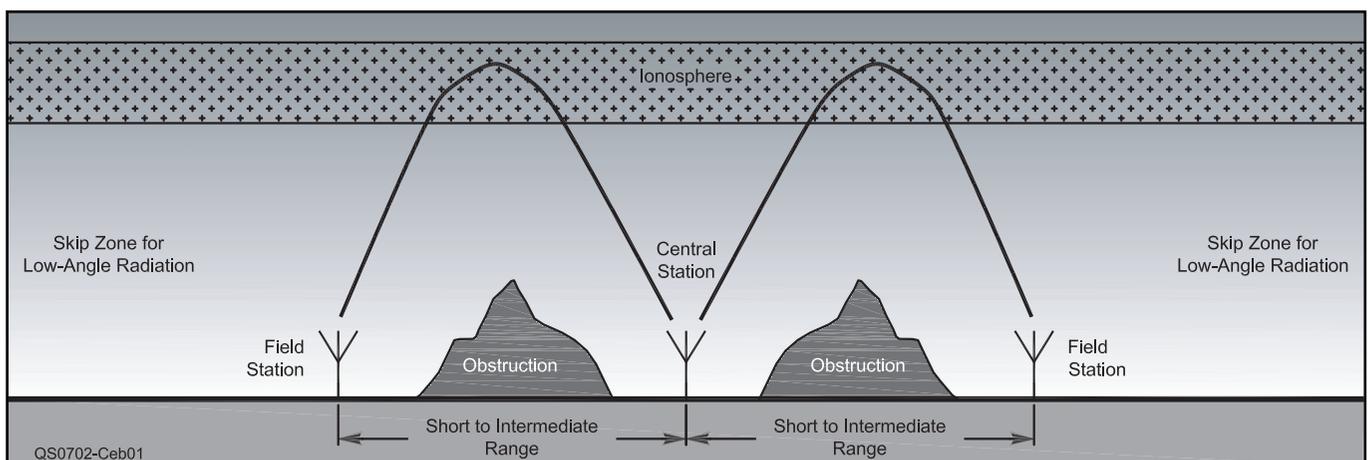
Figure 1 — The NVIS situation.

will be 3.9 MHz. A wavelength is about 252.2 feet at this frequency. The tabular data will be in fractions of a wavelength, so this number is handy for translating the information into numbers for physical planning. The trends that we uncover will be applicable throughout the lower HF range. Figure 3 sketches the four-dipole configurations that we shall examine.

The first case of a dipole over bare ground has two goals. One aim is to see at what antenna height we obtain maximum upward gain. The second purpose is to put to rest a certain persistent myth about NVIS dipoles, namely, that a super low height provides a gain advantage. Table 1 provides expanded information on the performance of a dipole over bare ground at heights ranging from

0.05 λ (about 12.5 feet at 3.9 MHz) up to a quarter wavelength (63 feet). The table provides gain values from *NEC-4* using the Sommerfeld-Norton ground calculation system (referred to as "high accuracy" in *EZNEC*). It also provides gain numbers reported by the only modeling program readily available during the early days of NVIS antenna analysis in the late 1980s and very early 1990s. That
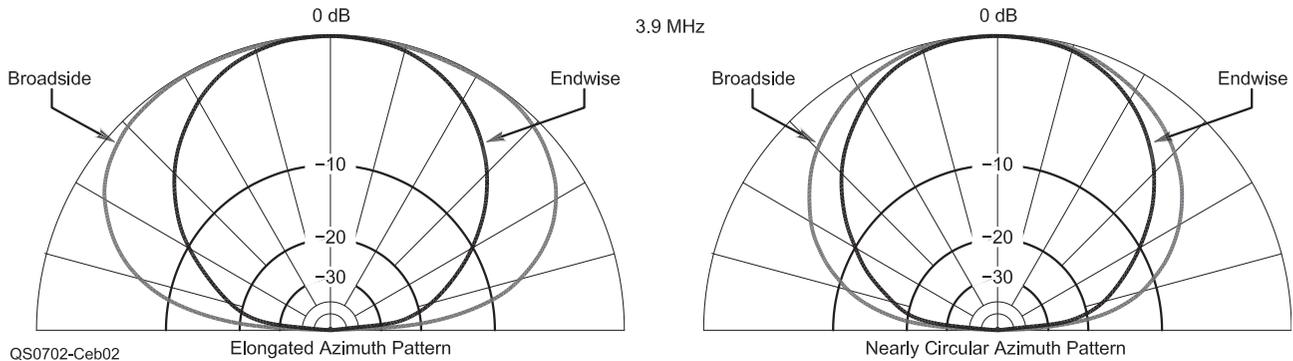
Figure 2 — Examples of more circular and more elongated NVIS azimuth patterns inferred from using elevation patterns and beamwidth values broadside to antenna wires and endwise to antenna wires.

Table 2
Modeled Performance of an NVIS Dipole at Various Heights Over Various Lower Structures at Ground Level (See Figure 3)

A. Single ½ λ Wire at Ground Level

Height (λ)	Gain (dBi)	Beamwidth (Degrees) Broadside	Endwise	Δ Gain (Versus Bare Ground)
0.125	6.21	105	65	0.41
0.15	6.50	108	65	0.27
0.175	6.59	112	67	0.19
0.2	6.53	118	68	0.14
0.225	6.35	124	70	0.10

B. 1 λ by 1 λ Screen

Height (λ)	Gain (dBi)	Beamwidth (Degrees) Broadside	Endwise	Δ Gain (Versus Bare Ground)	Δ Gain (Versus Single Wire at Ground)
0.125	6.78	100	65	0.98	0.57
0.15	7.03	104	65	0.80	0.53
0.175	7.09	108	66	0.69	0.50
0.2	7.01	113	67	0.62	0.48
0.225	6.82	119	70	0.57	0.47

C. Nine ½-λ wires spaced 0.1 λ

Height (λ)	Gain (dBi)	Beamwidth (Degrees) Broadside	Endwise	Δ Gain Versus Screen
0.125	6.80	101	66	0.02
0.15	7.00	105	66	−0.03
0.175	7.03	109	67	−0.06
0.2	6.92	115	69	−0.08
0.225	6.72	120	71	−0.10

Notes
1. All lower structures are 0.001 λ above average ground to permit model to run in both NEC-2 and NEC-4.
2. Single wire is AWG no. 14 copper.
3. Screen consists of 0.1 λ by 0.1 λ cells using 1 inch wire 0.001 λ above ground. This structure does not fully simulate a solid surface, which would increase gain values slightly. However, it may be reasonably accurate to typical amateur screen materials, such as chicken wire.
4. Nine-wire system is 0.8- λ long, broadside to the dipole and uses AWG no. 14 copper wires 0.001 λ above ground.

program was *MININEC*. As early as February, 1991, Roy Lewallen, W7EL, provided warnings to *QST* readers about the limitations of the *MININEC* simplified ground calculations system in his article "*MININEC*: The Other Edge of the Sword." Unfortunately, even today, many beginning modelers do not heed the warning. As the table shows, when we place a horizontal antenna below about 0.2 λ, *MININEC* reports an ever-inflating gain value. At the lowest height used in the table, the actual gain is 8 dB lower than the *MININEC* report. See the Straw article for the safety concerns and the supposed noise advantage of very low antennas.

The table shows a gain peak with the antenna about 0.175 λ above ground. Although this height will be consistently the peak gain height for all of our simple antennas, heights from about 0.125 λ up to about 0.225 λ are perfectly acceptable. As we raise the antenna in small increments, we notice a slow rise in the endwise beamwidth, but a more rapid rise in the broadside beamwidth. Hence, we can go some way toward tailoring the circularity or elongation of the pattern simply by varying the height without seriously subtracting from the available gain.

The remaining 3 configurations for an NVIS dipole reflect methods that some operators use or should use to improve performance. Table 2 supplies the corresponding modeling data, but restricts the height range to values from 0.125 λ to 0.225 λ. The first supplemented dipole uses a single ½ λ wire at ground level below the dipole. This wire and all other antenna supplements use modeled heights of 0.001 λ above ground so that the available models will run on both *NEC-2* and *NEC-4*.[1] The installer's goal is to create

[1] Models for the antennas discussed in this "Antenna Options" column are available in *EZNEC* format at the ARRL Web site. Go to **www.arrl.org/qexfiles** and look for 1x06_AO.zip.

a virtual Yagi pointed upward. Contrary to expectations, the table shows a very limited improvement in maximum gain, with the best improvement at the lowest height. Ground reflections do not occur just below the antenna wire, but over a very wide area in all directions from the antenna.

Studies of HF dipole arrays used for short-wave broadcasting and of VHF/UHF planar reflector arrays show that a conductive screen forms a very useful reflecting surface based on principles derived from optics. Such screens perform best when they extend at least ½ λ beyond the driven elements in all directions. The model with the screen in Figure 3 and in Table 2 uses a screen that is

1 λ by 1 λ on a side. The cells are 0.1 λ on a side. To simulate a solid screen, the wire would have to be very thick and using that wire would prevent the screen from sitting 0.001 λ above ground. So I reduced the wire size to a 1 inch diameter. The reduced wire size leaves the screen full of holes, and reduces the reported gain. However, it may also better reflect the likely amateur use of inexpensive materials such as chicken wire in which the junctions are not durably connected. The tabulated data shows a nearly constant improvement over the dipole and single-wire reflector. It also shows a decreasing improvement over the dipole above bare average ground. Still, the peak gain height
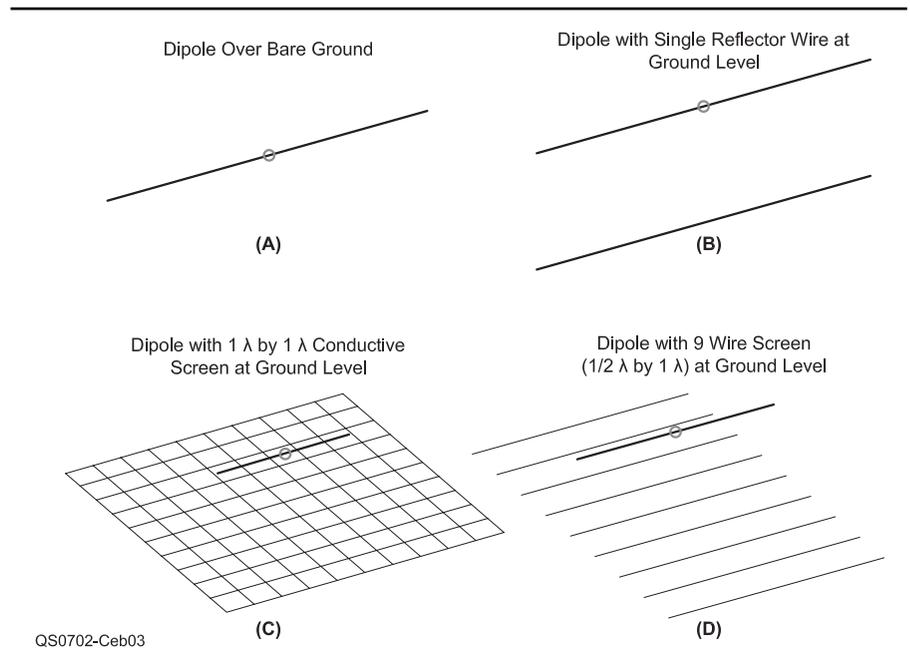
Figure 3 — Four NVIS linear dipole configurations. See Tables 1 and 2.

Table 3
Modeled Performance of Dipoles at 0.175 λ Height Over Various Ground Qualities

A. Dipole Above Bare Ground

Ground Quality	Gain (dBi)	Beamwidth (Degrees)		Conductivity (S/m)	Relative Permittivity
		Broadside	Endwise		
Very Poor	6.21	122	67	0.001	5
Average	6.40	113	66	0.005	13
Very Good	7.39	107.4	66	0.0303	20
Perfect	8.19	103	65	—	—

B. Dipole Above 1 λ by 1 λ Screen

Ground Quality	Gain (dBi)	Beamwidth (Degrees)		Conductivity (S/m)	Relative Permittivity	Δ Gain
		Broadside	Endwise			Versus Bare Ground
Very Poor	6.88	108	65	0.001	5	1.82
Average	7.09	108	66	0.005	13	0.69
Very Good	7.56	106	66	0.0303	20	0.17
Perfect	8.19	103	65	—	—	0.00

remains at 0.175 λ. The broadside beamwidth shows a 4° to 5° reduction with the screen in place, but the endwise beamwidth does not change at all.

We can simulate a full screen with a series of wires at ground level if we use enough of them. The final configuration uses nine AWG no. 14 copper wires at a height of 0.001 λ. The wire spacing is 0.1 λ. Each wire is only slightly longer than the dipole itself. With the 9 wires forming a field that is 0.8 λ long, the final section of Table 2 shows performance virtually identical to the performance with a full screen. Smaller numbers of wires or total field sizes produce lesser performance levels. (Since the single reflector proved so ineffective and since the screen is simpler to model than the nine wires, analyses of other simple antenna will contrast bare-ground and screen performance. The nine wire field, however, is always available as an alternative to a screen and may be easier to install.)

Let's pause here to look at an important side question: How does ground quality affect the improvement level offered by the screen or the nine-wire field? I modeled the dipole at a height of 0.175 λ over bare ground and over the screen using several ground quality levels, all of which appear in Table 3. The worse the soil quality, the greater the improvement offered by the ground-level screen. Over very poor soil, the gain improvement is nearly 2 dB, but over very good soil, the improvement drops to only 0.2 dB. In no case of solid ground does the use of a screen seriously approach the level of a perfect ground, although at sea, one might come very close. The conclusion is that NVIS antennas over poorer grades of soil may benefit significantly from a screen or a nine wire reflector.

The Inverted-V

Testing NVIS inverted-V antennas adds another variable to our modeling efforts. Let's assume that we use a fixed height for the ends of the V. For safety, I placed the ends 0.05 λ (about 12.5 feet at 3.9 MHz) above ground. As I surveyed changing top heights, I restored the antenna to near resonance, which drew the ends in toward the center and increased the angle of the wire relative to the ground. Figure 4 shows the bare ground and screen configurations. Table 4 presents the test results, including the wire angle.

The inverted-V over bare ground shows much less gain than a dipole. The effective height of the entire wire is about ⅔ the distance between the lower and the upper ends. Hence, the peak gain is approximately the same as the dipole over bare ground at a height between 0.1 λ and 0.125 λ. Consistent with the dipole models, the peak gain for the inverted-V occurs at a top height of 0.175 λ.

One advantage of an inverted-V is that the endwise beamwidth increases by from 10° to 20° relative to the dipole. In fact, the higher that we place the V, the more circular the pattern becomes. At a height of 0.225 λ, we find only a 23° difference between the broadside and the endwise beamwidth reports. The wire angle at this height is about 46°, a value that we cannot safely achieve at lower top heights. Obviously, the greater the wire angle relative to ground, the more omni-directional the pattern becomes.

For any given top height, the effective height of an inverted-V will be lower than a linear dipole at the same height. As a consequence, the inverted-V tends to benefit more from the presence of a ground-level screen (or its nine-wire substitute). The lower part of Table 4 shows a 1-dB or greater improvement in gain. As well, it shows a slight improvement in the circularity of the azimuth cover-age due to a small shrinkage in the broadside beamwidth.

One strategy for setting up an NVIS antenna system is to use a pair of inverted-V antennas — one for 80/75 meters, the other for 40 meters — using a common center support and a common feed point. If the antennas are at right angles to each other, interactions between the two sets of wires will be minimal. The limitation of such a system is that we end up with both bands using heights that are not optimal. 0.125 λ at 3.9 MHz is close to 0.23 λ at 7.2 MHz. While both heights fall within the scanned range for our test cases, one or the other may yield a pattern shape that is not ideal. We may use the same center support for both bands, but setting up separate antennas optimized for the best height and wire angle (over a reflector screen) may

Inverted-V Over Bare Ground

Inverted-V Over 1 λ by 1 λ
Screen at Ground Level

(A)

(B)

QS0702-Ceb04
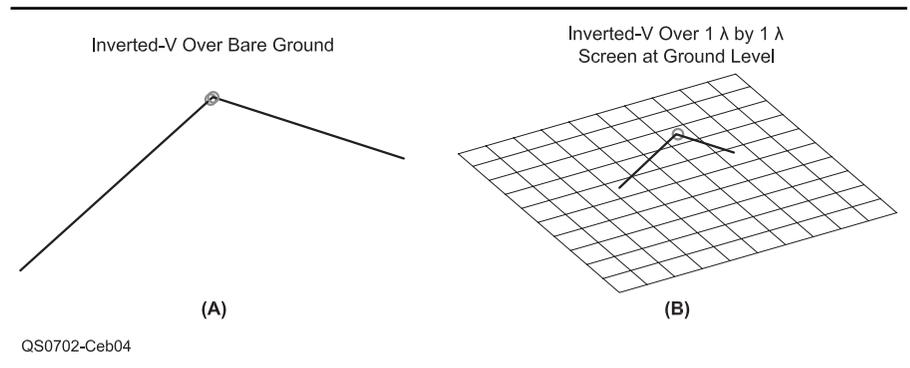
Figure 4 — NVIS inverted-V configurations. See Table 4.

Table 4
Modeled Performance of an Inverted-V at Various Heights Over Various Lower Structures at Ground Level (See Figure 4)

A. Inverted-V above bare ground

Height (λ)	Gain (dBi)	Beamwidth (Degrees)		Wire Angle (Degrees)
		Broadside	Endwise	Versus Ground Level
0.125	3.88	102	77	18
0.15	4.18	104	79	26
0.175	4.32	106	82	31
0.2	4.31	109	86	38
0.225	4.17	113	90	46

B. Inverted V Above 1 λ by 1 λ Screen

Height (λ)	Gain (dBi)	Beamwidth (Degrees)		Δ Gain
		Broadside	Endwise	Versus Bare Ground
0.125	5.24	98	77	1.36
0.15	5.40	100	79	1.22
0.175	5.42	102	82	1.10
0.2	5.32	105	86	1.01
0.225	5.11	109	91	0.94

Notes
1. All inverted Vs use AWG no. 14 copper wire above average ground.
2. V ends are fixed at 0.05 λ above ground. Lengths adjusted as height increases to establish near resonance.

3. Screen consists of 0.1 λ by 0.1 λ cells using 1 inch wire 0.001 λ above ground.

let us achieve near circularity of coverage or just the degree of pattern elongation that we need for the intended coverage area.

The 1 λ Loop

An overlooked antenna for NVIS work is the 1 λ loop. Each side of the loop is only about half the length of a dipole for the same frequency. If we plan to supplement the antenna with a screen or other reflection means, the loop may prove to be more compact than a dipole or a V with a screen below. As well, we can nest loops for each lower HF band that we wish to cover. Figure 5 shows the bare-ground and the screened configurations, the data for which appear in Table 5.

Compared to a dipole, the 1 λ loop provides slightly higher maximum gain levels and slightly more circular patterns. At the height of maximum gain, the loop pattern is about 24° less oval than the dipole pattern, as measured by the difference between the broadside and the endwise beamwidth values. For the loop, the broadside direction passes through the mid-side feed point and the midpoint of the opposite side. The endwise pattern passes through the opposing two sides without a feed point. The use of a 1 λ by 1 λ screen below the loop at ground level provides slightly less added gain than it does for a dipole.

In some respects, the 1 λ loop provides the best of the dipole and the inverted-V worlds. It has the gain of the dipole and the nearly circular pattern of the V. It does require four corner supports, however, and the feed point is well above the range for a good match to common 50 Ω coaxial cable. The latter problem disappears if we add a ¼ λ section of 70 to 75 Ω cable.

High-Gain NVIS Arrays

The simple antennas that we have explored offer a balance between gain and beamwidth. Since we are dealing with nearly circular azimuth patterns with only one main lobe, the only way to increase upward or maximum gain is to decrease the beamwidth in one or both directions. Hence, high-gain arrays are not necessarily for everyone. The operator who needs to use elevation angles down to say 45° may obtain better results with one of the simple antennas. However, a central station that requires only short-distance communication may find an advantage in concentrating his or her signal upward. In fact, by using a common wire array configuration, we may obtain up to 6-dB additional gain.

Perhaps the two most common NVIS arrays used to increase upward gain are the "Jamaica" and the "Shirley" array. Actually, both arrays are forms of the lazy-H facing the sky. Moreover, for raw upward gain,

Figure 6 — NVIS Shirley array outline and elevation patterns. See Table 6.
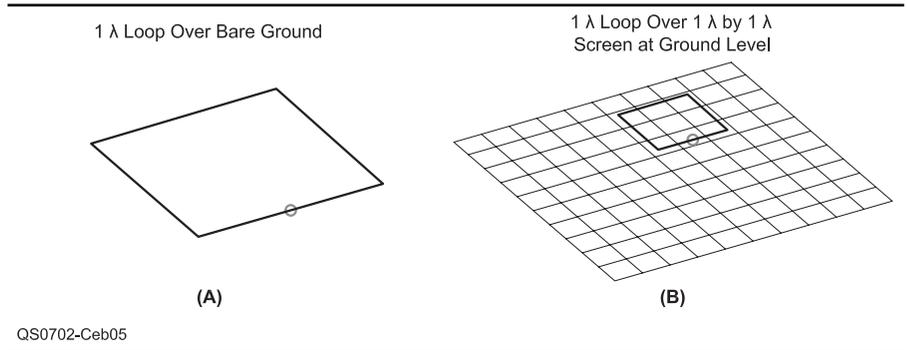
Figure 5 — NVIS 1 λ loop configurations. See Table 5.

Table 5
Modeled Performance of a 1-λ Loop at Various Heights Over Various Lower Structures at Ground Level (See Figure 5)

A. 1 λ Loop Above Bare Ground

Height (λ)	Gain (dBi)	Beamwidth (Degrees)	
		Broadside	*Endwise*
0.125	6.45	84	69
0.15	6.85	87	69
0.175	7.02	90	70
0.2	7.03	95	72
0.225	6.89	101	74
0.25	6.64	108	78

B. 1 λ Loop Above 1 λ by 1 λ screen

Height (λ)	Gain (dBi)	Beamwidth (Degrees)		Δ Gain
		Broadside	*Endwise*	*Versus Bare Ground*
0.125	7.34	91	68	0.89
0.15	7.58	84	69	0.73
0.175	7.64	87	70	0.62
0.2	7.58	91	71	0.55
0.225	7.41	97	74	0.52
0.25	7.14	103	77	0.50

Notes
1. All 1 λ loops use AWG no. 14 copper wire above average ground.

2. Screen consists of 0.1 λ by 0.1 λ cells using 1 inch wire 0.001 λ above ground.
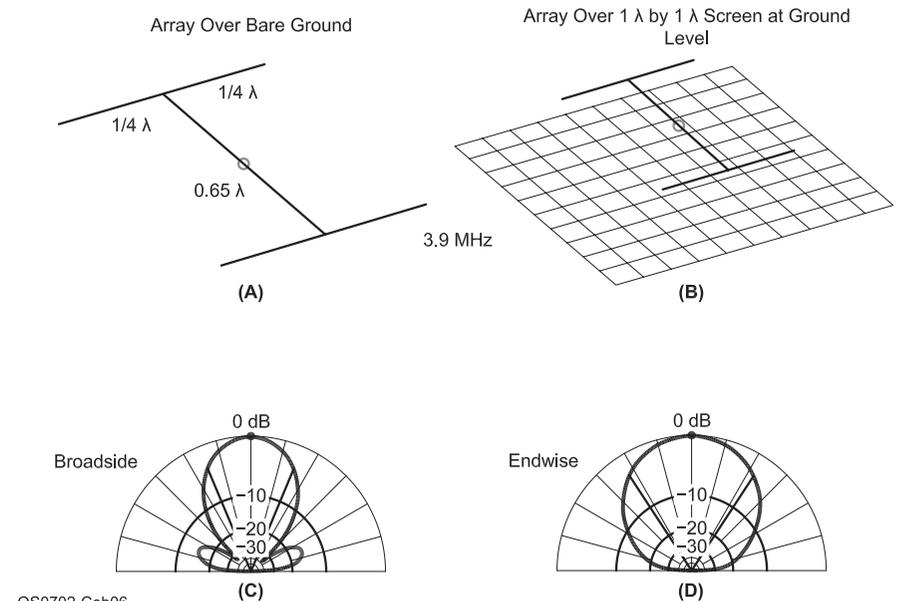
these arrays overlook the best of the lot: the extended (or expanded) lazy-H. Let's quickly sample all three antennas, both with and without a screen (or multi-wire) supplement. In each case, the antenna itself will use the same AWG no. 14 copper wire common to the simple antennas. We feed each antenna element in phase with equal length lines to a central feed point. The feed point impedance will vary according to the element length, the spacing, and the characteristic impedance of the phasing lines. Table 6 provides the modeled data for all three antennas.

The Shirley array, shown in Figure 6 uses ½ λ elements spaced about 0.65 λ apart. Some versions use folded dipole elements for a presumed match to the phase lines, but that aspect of the antenna construction plays no role in establishing the basic gain and pattern data. At a height between 0.175 λ and 0.2 λ, the antenna shows a little over 4-dB gain over a dipole. The price that we pay for the gain is a very significant reduction in the broadside beamwidth (by nearly 70°), but not in the endwise beamwidth. See the lower part of Figure 6. The result is a highly elongated oval that favors the directions off the ends of the elements. The use of a 1 λ by 1 λ screen adds nearly a dB to the gain without altering the beamwidth values.

The Jamaica array, shown in Figure 7, uses a standard lazy-H configuration: two 1-λ elements with a ½-λ space between them. Over bare ground, it improves upward gain by a full dB over the Shirley array, but the addition of a 1.5 λ by 1 λ screen adds less than a half dB more. In both cases, the Jamaica height is the same as the Shirley height. One way to view the circularity of the patterns is to take the ratio of the broadside to endwise beamwidth values. The dipole over bare ground shows a broadside-to-endwise ratio of 1.7:1. In contrast, the Jamaica array has a ratio of less than 1.3:1 over bare ground, as illustrated by the patterns in the lower half of Figure 7. Both numbers are drawn from the height of maximum gain. (In contrast, the Shirley array showed a broadside-to-endwise ratio of 0.6:1.) Note that the use of collinear half-wavelength elements and half-wavelength spacing yields no sidelobes. The wider spacing of the Shirley elements revealed the emergence of low-angle broadside lobes.

For raw gain, we can do little better than increase the element length to 1.25 λ and use a 0.65 λ space between the elements. Figure 8 shows the results, along with patterns that reveal the emergence of sidelobes in both the broadside and endwise directions. The user will have to determine whether the lower-angle sidelobes present a danger of increased noise pick-up based on the level and types of noise sources for the given location.

If the sidelobes do not pose a problems, then the extended lazy-H array adds nearly

2 dB to the gain offered by the Jamaica beam, without a screen or with the requisite 2 λ by 1 λ screen needed by the extended lazy-H. (Of course, a field of wires about 1.5 λ long and extending about 0.4 to 0.5 λ beyond the broadside limits of the active array may substitute for the screen.)

The screen adds only about 0.4 dB gain to the bare ground version of the antenna. Essentially, the extended lazy-H configuration provides 12.5 to 13 dBi maximum gain

over average ground, compared to 6.4 to 7.1 dB for a dipole at roughly the same height. In exchange, as shown by the lower part of Figure 8, we further narrow the beamwidth — down to 44° broadside and 31° endwise. The 1.4:1 ratio shows fairly good azimuth circularity.

Assuming that we can handle the Lazy-H array feed point impedance values with an antenna tuner, the extended version of the array offers a further unadvertised benefit that

---

Table 6

Modeled Performance of Several High-Gain NVIS Arrays Based on Various Forms of the lazy-H (See Figures 6, 7 and 8)

A. "Shirley" array (0.65 λ spaced 0.5 λ elements): Screen 1 λ x 1 λ

Environment	Gain (dBi)	Beamwidth (Degrees)		Δ Gain
		Broadside	Endwise	Over No Screen
No screen	9.90	44	68	
1 λ x 1 λ screen	10.88	45	66	0.98

B. "Jamaica" array (0.5 λ spaced 1 λ elements) : Screen 1.5 λ x 1 λ

Environment	Gain (dBi)	Beamwidth (Degrees)		Δ Gain
		Broadside	Endwise	Over No Screen
No screen	10.88	56	44	
1 λ x 1 λscreen	11.26	55	44	0.38

C. Extended lazy-H array (0.65 λ spaced 1.25 λ elements): Screen 2 λ x 1 λ

Environment	Gain (dBi)	Beamwidth (Degrees)		Δ Gain
		Broadside	Endwise	Over No Screen
No screen	12.63	44	31	
1 λ x 1 λ screen	13.03	44	31	0.40

Notes
1. All antenna arrays use AWG no. 14 copper wire above average ground.
2. Screens consist of 0.1 λ by 0.1 λ cells using 1 inch wire 0.001 λ above ground.
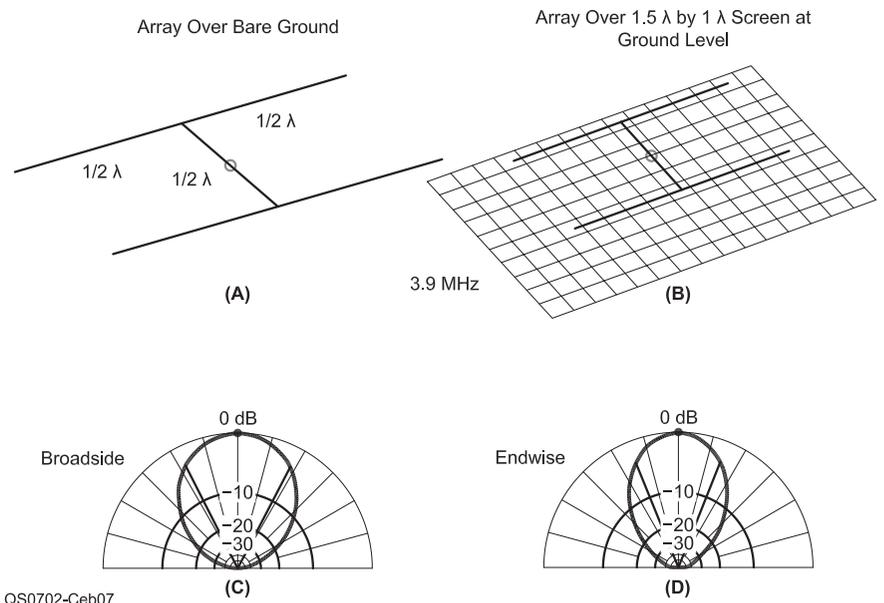
Figure 7 — NVIS Jamaica array outline and elevation patterns. See Table 6.
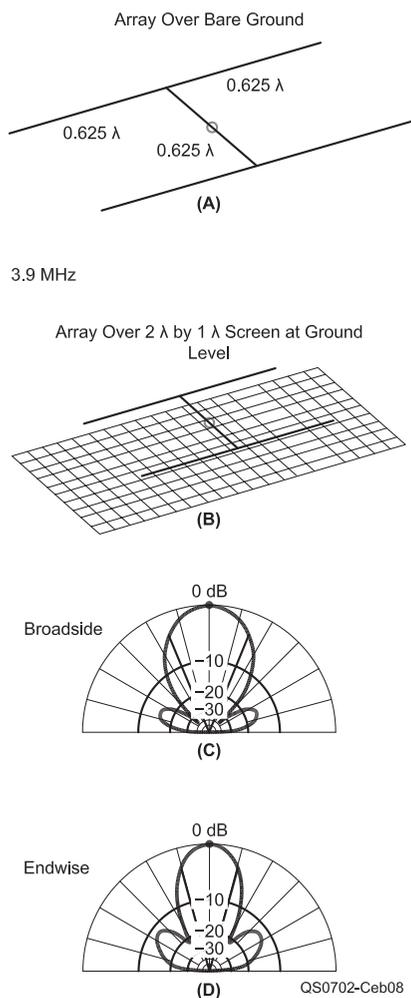
Figure 8 — NVIS extended lazy-H array outline and elevation patterns. See Table 6.

Figure 9 — Comparative broadside and endwise elevation patterns of a 1 λ loop and an extended lazy-H array over bare ground at the same height.

stems from its higher gain level. We can afford to place an extended lazy-H a bit higher than the optimal height and set the element length at 1.25 λ on 7.2 MHz, with a 40 meter spacing of about 0.6 λ to 0.65 λ. The gain deficit relative to an optimal height will be small. At 3.9 MHz, the array will be a little over half as high, and the element lengths and the spacing will be half as much. The array will still perform well, with a gain level intermediate between a dipole and a full extended lazy-H. At the lower frequency, we would find no sidelobes.

The lure of additional gain often blinds us to other considerations that may affect our operation. Throughout these notes, I have tried to give equal strength to gain and beamwidth comparisons. Which factor requires greater weight in deciding on an NVIS antenna requires an operator decision. If operations require more than short range, then the added gain of the lazy-H configurations may not be

an advantage. Then Figure 9 may be of interest. It shows overlaid patterns for a 1 λ loop and for the extended lazy-H array, both over bare ground. The increased gain potential of the lazy-H captures our initial attention. If we count upward to the 45° elevation angle, however, we discover that the loop has a significantly higher gain in that elevation direction. In fact, at that angle, the lazy-H has almost no gain, since it is the angle for the null between the main lobe and the sidelobe in both the broadside and the endwise patterns.

Conclusion

We have looked at a variety of central-station NVIS antennas for amateur use with an eye toward finding the optimal height of maximum gain (between 0.175 λ and 0.2 λ). We also examined the level of pattern circularity achieved by these simple designs. We also explored a few high-gain NVIS arrays, as well as the level of benefit offered by ground screens or multi-wire substitutes. Which antenna might be correct for you de-

pends on your available space, your access to supports, and also the type of operating that you do. For casual chats with a neighbor who lives beyond yon hill (or even beyond the hill beyond yon hill), a high gain array may be a nearly ideal NVIS antenna.

For the emergency operator, high gain alone may not solve all challenges, especially if there is a need to reach beyond short range into the intermediate range that still falls within the skip zone. One of the simpler antennas may better serve the requirement. In his article, Straw noted the need for multiple relays to route important messages outward and then back inward toward targets, many of which required NVIS-type propagation. Under these conditions, the right antenna — abetted by high operator skill and experience — proved invaluable. In fact, it gave a contemporary rationale for continuing to call our national organization the American Radio *Relay* League. Indeed, an operator who copies precisely and relays accurately is as important as the antenna that he or she uses. QEX

Letters to the Editor

RF Power Amplifier Output Impedance Revisited (Jan/Feb 2005; "Letters" Jul/Aug 2006)

Doug,

This concerns your response to John Belrose's letter to the editor, to clarify three issues: 1) use of a reverse-power generator to determine output source resistance R of RF PAs, 2) conjugate match with RF power amplifiers, and 3) nondissipative output resistance of RF PAs.

1. Mr. Rauch used a reverse power generator, RPG, to determine the value of the output source resistance R of the RF PA.[1] The frequency of the RPG was chosen [to be] as close as possible to that of the PA to permit reading only the RPG signal at the Ts. Since the source resistance R is nondissipative (see below) when $R = Z_0 = 50 \, \Omega$, this 50-Ω-sourced probe signal is transferred back into the RF PA's anode circuit, where it modulates the PA transmit signal.

Certainly, the experiment works. Rauch found that when output source resistance of the PA (at the output terminals) equals the Z_0 of the line, voltages at all three Ts are equal. However, when $R > Z_0$ the voltage at the T nearest the output terminals is greater than at the other Ts, showing a standing wave established at the RPG frequency; conversely, when $R < Z_0$ the voltage at the nearest T is less than that at the others. Consequently, adjustment of source resistance R to achieve equal voltages at all Ts results in no standing wave, and indicates $R = Z_0$. Using the IEEE load-variation method, Maxwell also measured the output to determine the value of the output source resistance of PAs.[2, 3, 4]

2. It is well known that a conjugate match (CM) exists if power delivery decreases with any change in load impedance. Thus a CM exists when the source and load [impedances] are either equal or conjugately related. Maxwell proved this experimentally several years ago, but the results were not published. Rauch also proved it by first adjusting the pi-network of the PA to deliver all available power into a reactive load, $50 + j50 \, \Omega$, measuring the resulting input impedance Z_{IN} of the network with the PA powered down, connecting an equivalent impedance Z_{IN} across the network input, and then measuring the impedance looking back into the network and finding impedance $Z_{OUT} = 50 - j50 \, \Omega$. (See Note 1.) Maxwell later performed a similar experiment, obtaining similar results.[4]

3. Maxwell's references[2, 3, 4] present proof that the source resistance of the RF PA appearing at the output terminals is nondissipative. It must be understood that the total

plate resistance in a PA tube comprises two separate resistances, one dissipative (R_{PD}) and the other nondissipative (R_L). R_{PD} is cathode-to-plate resistance, responsible for heating the plate due to the electron bombardment at the plate. The power dissipated is equal to the product of the instantaneous voltage and current flowing between the cathode and plate, which is irrelevant to plate resistance R_P, a virtual, nondissipative resistance. On the other hand, R_L is established by the voltage-current relationship at the input of the pi-network, a ratio R/L, which cannot dissipate power. Consequently, the output source resistance of the PA is the ratio of the current to voltage appearing at the output terminals of the network, indeed, a nondissipative resistance.

— *73, ARRL TAs Walter Maxwell, W2DU,* **walt@w2du.com;** *and John (Jack) Belrose, VE2CV,* **john.belrose@crc.ca**

Notes

[1] J.S.Belrose, W. Maxwell and C.T. Rauch, "Source Impedance of HF Tuned Power Amplifiers and the Conjugate Match," *Communications Quarterly*, Fall 1997, pp 25-40.

[2] Walter Maxwell, W2DU, "On the Nature of the Source of Power in Class B and C RF Amplifiers," *QEX*, May/Jun 2001, pp 32-44.

[3] Same as above, but also appearing in *Reflections II*, published by Worldradio Books.

[4] Walter Maxwell, W2DU, "Additional Proof of Conjugate Match and Non-Dissipative Source Resistance In RF Power Amplifiers," *Reflections 3*, (in publication, not yet released, but can be accessed at **www.w2du.com**, click on 'Preview of Reflection 3 Articles, then click on Chapter 19A.)

Gentlemen,

Thank you, but you've not addressed my basic objection: If an amplifier's source resistance were not ohmic, then it would be measured as a pure reactance in a measurement of s22 (output reflection coefficient). Many device manufacturers report s22 as the conjugate of the optimal load impedance, which is not necessarily what you observe when probing with an RPG.

My theory is that resistance is the defining characteristic of any process that converts energy from one form to another and conveys it. Even the resistances of free space ($120\pi \, \Omega$) and of $50 \, \Omega$ coaxial cable embody such processes. But if no RPG energy were reflected by the source (s22 = 0), that would reveal an ohmic resistance. Were s22 = 1, you could state that the match were optimal, in that it maximized energy transfer, but not that it were conjugate. That distinction is made in many relevant texts.

— *73, Doug Smith, KF6DX,* QEX *Editor;* **kf6dx@arrl.org**

Doug,

I find no fault in Tom Rauch's instrumentation except that his test signal may be too high to avoid disturbing the R_S produced by the tube alone. It must be small enough, when stepped up by the pi network, to be much smaller than the difference between the peak of the RF plate voltage and the dc screen voltage of a tetrode.

In a private latter, Tom explained in more detail exactly how he tuned and loaded to achieve zero reflected power of the test signal. Clearly, he was operating the tubes well into the nonlinear plate-saturation region where his [$R_S = R_L$] result can be expected. Tom never claimed that his "zero-reflection" operating point was within an acceptable linear operating range.

Little has been said or written about how R_S varies with signal level in real tubes. Typically, R_S is very high when the signal level is small, and drops by a factor such as 3 to 10 times when the signal reaches the limit of acceptable linearity. Then, as the drive is increased further, the value of R_S starts to drop rapidly until it passes through the point where $R_S = R_L$ and, with still more drive, R_S can become much lower than R_L. The tube will operate fine if you back off the drive enough to get acceptable IM performance. You may have to make several tries at different drive levels to find the one that, when the drive is backed down enough, will give the desired RF power output.

About 50 years ago, I conceived a system to automatically tune an SSB transmitter. The system did not look for a peak or dip in anything. What is more, it tuned at a low power level where R_S was nowhere near R_L. I designed an automatically tuned 45-kW PEP HF transmitter that met a certain military specification of 46 dB PEP/IMD3 ratio. It would tune from any frequency in the 2 to 30 MHz range to any other in 5 to 15 seconds. The plate was tuned for a 180° phase difference from the RF grid voltage. Plate loading was adjusted to seek the design value of grid-to-plate RF voltage gain.

I also designed a 10-kW PEP power amplifier that was tuned in the same way, except that an operator turned the knobs to zero the meters. The Collins 30S-1 and 30L-1 linear amplifiers use the same scheme to get the correct loading adjustment. It distresses me that today, it is a rare piece of ham equipment that comes anywhere near the IMD performance that Collins Radio met back then.

— *Sincerely and 73, Warren Bruene, W5OLY, 7805 Chattington Drive, Dallas, TX 75248-5307.*

Q Calculations of L-C Circuits and Transmission Lines: A Unified Approach (Sep/Oct 2006)

Dear Doug,

On p 43, middle column, 1st paragraph after Eq 1, the first sentence is: "The simple series model $R_S - jX...$" This should be, of course, "$R_S + jX$", with X carrying the sign of either inductive reactance, X_L (+), or capacitive reactance, X_C (−).

By the way, the *ARRL Handbook*, 2000 edition, the latest that I have, on p 6.13, Eq 31 gives $X_C = 1/2\pi fC$, which is incorrect. The book *Physics for Dummies* gives it correctly with a negative sign, $X_C = -1/2\pi fC$. I hope this is presented correctly in the latest editions of the *Handbook* or will be corrected for the 2007 edition.

— *Regards, Larry Joy, WN8P, ARRL LM;* **ljoy@kantronics.com**

Hi Larry,

You're right, of course. In the 2007 edition of the *Handbook*, the first equation in which the error appears is Eq 53 on p 4-24 of Chapter 4. It propagates through the following discussions for several pages. We will call this to the attention of the *Handbook* editor, so the 2008 edition can reflect that and other significant corrections regarding electromagnetism. Because of you, some of us finally got around to reading those chapters in the *Handbook* that EEs usually ignore!

— *73, Doug, KF6DX.*

Dear Doug,

Thank you for the excellent article.

I downloaded the author's *MathCAD* files (TRL_Q_Calc1.mcd, TRL_Q_Calc-PCB1.mcd, and Monopole-Ralph.mcd) from the ARRL *QEX* Web site. I was disappointed, however, to find that my older version (6+) of *MathCAD* could not open the files. Is there some way that I can receive a printout of these files, either hard copy or PDF format?

May I suggest that *QEX* and *QST* make available for download PDF copies of *MathCAD* documents along with the .mcd files themselves?

— *73, Don Shoop, NØASG;* **tazzoc23@mindspring.com**

Don,

Thanks for calling this to my attention. I have now sent copies of the files that will be compatible with *MathCAD* V6, and the 9x06_Audet.zip file posted on the *QEX* files section of the ARRL Web site (**www.arrl.org/qexfiles**) has been updated to include those files. Also, it seems that *MathCAD* V6 cannot display the 3D graphs. I have supplied a version of the graphs as a PDF file, also included in the file for download.

— *73, Jacques Audet, VE2AZX;* **jaudet@hotmail.com**

In Search of New Receiver Performance Paradigms, Part 1 (Nov/Dec 2006)

Hi Doug,

[Those are some] nice articles on receiver measurement. A few comments:

UK marine receiver specifications even 40 years ago required you to meet the SNR on top of any spur. For channelised receivers, you chose your channels very carefully!

It would seem that for testing the new breed of receiver (and the method is applicable to the conventional type), we could draw lessons from the old FDM [frequency-division multiplexed] telephone systems. There, IMD testing was done by producing wideband noise with a notch in it at the receive frequency, and measuring the amount of noise received. For radio, you would want to add a wanted signal so that you could measure the degradation in SINAD for analogue or BER for digital signals. Such a test would need a crystal notch filter (or a very sophisticated generator) but would be very sensitive, as it would pick up "intermod," phase-noise second-order AM effects, cross modulation and spurious responses.

I'm afraid that you're a bit out-of-date on European regulation. We don't necessarily have to do third-party testing. The Radio & Telecommunications Terminals Directive allows a manufacturer to self-certify equipment against a harmonised standard, like one that has been published as such in the Official Journal of the EU. The applicable standards for amateur equipments are EN301 489, Parts 1 and 15. Those are EMC standards, since there are no product standards for amateur equipment. That situation is unlike commercial equipment, which has a product standard and an EMC standard. The Directive wasn't too popular with test houses, of course, since they lost a lot of business.

The standards may be obtained free at **www.etsi.org**. One interesting point is that the European Commission are very much against having receiver requirements in harmonised standards, unless they are necessary for purposes such as ensuring that a network operates correctly. Interestingly, even prior to that, there was little (if any) notice taken of the IEC documents on receiver testing. I think part of this is the way that ISO works, which is on a national delegation basis, rather than on an industry basis.

Finally, the homebrew equipment of radio amateurs (as defined by ITU), and Amateur Radio kits, are exempt from the EU regulations. That's thanks to some good work by IARU Region 1 some 20 years or so ago. Mssr. Gaston Bertels, ON4WF, did a lot of that necessary work.

— *73, Peter Chadwick, G3RZP, Chairman, ETSI (European Telecommunications Standards Institute) Task Group ERM_TG30;* **peter.chadwick@Zarlink.Com**

Hi Peter,

First off: Congratulations on your appointment to the Task Group! Second: I knew I could count on you to bring me up to date on EU regs — thanks. I believe many readers are, like me, unaware of those changes in the requirements for the CE mark, just as they are ignorant of EU requirements for power-factor correction in power supplies.

According to ARRL, controversy has arisen about whether what I wrote is true: that commercially sold amateur gear is subject to the same Part-15 regulations here as is everything else (*Minutes of the Executive Committee*, No. 480, Oct. 20, 2006, **http://www.arrl.org/announce/ec_minutes_480.html**). It will be interesting to see what the FCC says about that issue, although it seems perfectly clear to me from the wording of the law, which is now quite old.

For my money, the method of determining receiver IMD performance must be measured in a way that clever designs can't circumvent. The procedure you mention seems as good as, or better than, any, but paramount is: You have to measure noise floor under exactly the same conditions under which you measure IMD. I'll have more to write about that.

I hope this finds you well.

— *73, Doug, KF6DX.*

Twenty-Five Candles Burning Brightly ("Empirical Outlook," Nov/Dec 2006)

Dear Doug,

A happy birthday to *QEX*, the best technical journal indeed. All of you do a fine job. Carry on! And long live Amateur Radio.

— *Ray Cramet, F8CB;* **f8cb@wanadoo.fr**

Doug,

I read your editorial in the new issue and said to myself, "Wow, has it really been 25 years already?" But sure enough, I went to my collection and there is issue number 1, 25 years ago. Some of my magazine collections have had to be trimmed down or even given away entirely in the course of six or seven moves, but not *QEX*. I still have the entire set, and that will *not* change.

I remember some of the ups and downs over the years. Your time at the helm has been a good one. We've had steady publication and a consistently excellent stream of content. My compliments on a job well done.

As I'm looking back at some of the other magazines you mentioned, in particular the only one that aimed to do what *QEX* has done so well, *ham radio* magazine, I'm glad to say that *QEX* has filled the void it left very nicely. About the only things I miss are those amazing cover drawings by Hans Evers, PAØCX. No hope of those [returning], I suppose.

— *73, Paul Koning, NI1D;* **ni1d@arrl.net**

KENWOOD
Listen to the Future

KENWOOD SKYCOMMAND
TURN IT ON!

Kenwood SkyCommand has FCC approval.

Allows Global communication through remote operation on HF frequencies at home or in the field utilizing Kenwood's TS-2000 series transceivers.

Kenwood's TH-D7AG or TM-D700A required for remote use.

Perfect for use in hurricane or tornado zones, as well as Search and Rescue areas for Long Distance Communications when other normal modes of communications are out.

A great tool to monitor propagation while doing other things at home!

No cables or adaptors to fool with or buy!

No software or computer required!!

Step by step setup and programming taking only minutes.

Ease of use.

See you local dealer for details.