QEX
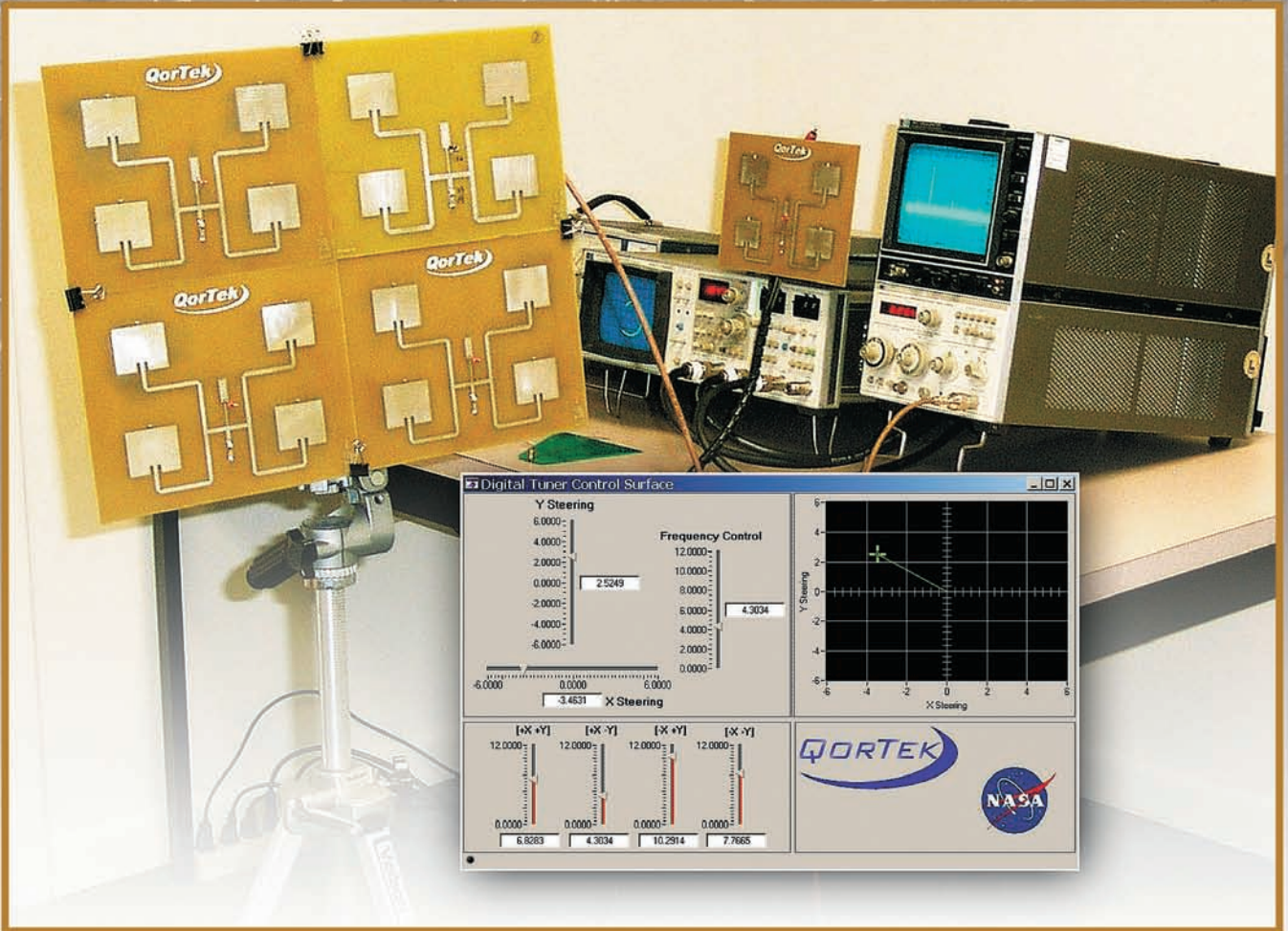
July/August 2006

A Forum for Communications Experimenters

N6TX models the nonlinear response of a NASA synthetic aperture radar antenna array, and controls the azimuth, elevation and frequency of the system with software.

CD-ROM Collections

2005 ARRL Periodicals CD-ROM

NEW EDITION! **ARRL Periodicals on CD-ROM** are fully-searchable collections of popular ARRL journals. Every word and photo published throughout the year is included!

SEARCH the full text of every article by entering titles, call signs, names—almost any word. SEE every word, photo (including color images), drawing and table in technical and general-interest features, columns and product reviews, plus all advertisements. PRINT what you see, or copy it into other applications.

System Requirements:

Microsoft *Windows*™. 1999, 2000, 2001, 2002, 2003, 2004 and 2005 editions support *Windows* and Macintosh systems, using the industry standard Adobe *Acrobat Reader*® (included).

Communications Quarterly CD-ROM

This CD-ROM collection covers volumes of *Communications Quarterly* published from 1990-1999. Gain access to advanced technical topics in articles which cover transmitter, receiver and transceiver projects, theory, antennas, troubleshooting and much more. High quality black-and-white page scans can be read on your computer screen or printed. Quickly **search** for articles by title and author, **select** specific year and issue, and **browse** individual articles and columns.
Requires Microsoft *Windows*™.
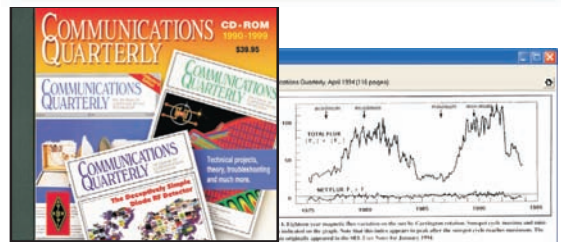
QST View

Every page of *QST*—all the ads, articles, columns and covers—has been scanned as a black-and-white image that can be viewed or printed.

BONUS! Latest, enhanced viewer—**AView**—included with any **QST View** purchase. **AView** is compatible with all sets of **QST View**.

ARRL CD-ROM Collections

ARRL Periodicals CD-ROM
(includes *QST*, *QEX*, *NCJ*)

#9574 Year 2005 .... **NEW!**	**$19.95**
#9396 Year 2004	**$19.95**
#9124 Year 2003	**$19.95**
#8802 Year 2002	**$19.95**
#8632 Year 2001	**$19.95**
#8209 Year 2000	**$19.95**
#7881 Year 1999	**$19.95**
#7377 Year 1998	**$19.95**
#6729 Year 1997	**$19.95**
#6109 Year 1996	**$19.95**
#5579 Year 1995	**$19.95**

QST View CD-ROM

#7008 Years 1915-29	**$39.95**
#6710 Years 1930-39	**$39.95**
#6648 Years 1940-49	**$39.95**
#6435 Years 1950-59	**$39.95**
#6443 Years 1960-64	**$39.95**
#6451 Years 1965-69	**$39.95**
#5781 Years 1970-74	**$39.95**
#5773 Years 1975-79	**$39.95**
#5765 Years 1980-84	**$39.95**
#5757 Years 1985-89	**$39.95**
#5749 Years 1990-94	**$39.95**
#8497 Years 1995-99	**$39.95**
#9418 Years 2000-2004	**$39.95**
#QSTV (all 13 sets) ... **Only $399** (Save $120)	

Communications Quarterly CD-ROM
#8780 (1990-1999) **$39.95**

Ham Radio Magazine CD-ROM*
#8381 Years 1968-76	**$59.95**
#8403 Years 1977-83	**$59.95**
#8411 Years 1984-90	**$59.95**
#HRCD (all three sets)	**$149.85**

QEX Collection CD-ROM
#7660 (1981-1998) **$39.95**

NCJ Collection CD-ROM
#7733 (1973-1998) **$39.95**

---

www.arrl.org/shop
1-888-277-5289 (US)

*Ham Radio CD-ROM, © 2001, American Radio Relay League, Inc. Ham Radio Magazine © 1968-1990, CQ Communications, Inc.

Shipping & Handling charges apply: US orders add $6 for one CD, plus $2 for each additional CD ($12 max.). International orders add $5.00 to US rate ($17.00 max.). Or, contact ARRL to locate a dealer. Sales Tax is required for orders shipped to CA, CT, VA, and Canada.

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

This NASA synthetic aperture radar antenna was modeled using the methods presented in "Adventures in Curve-Fitting." The graphical user interface display is the software that implements a nonlinear tuning algorithm.

Features

Columns

Jul/Aug 2006 *QEX* Advertising Index

Empirical Outlook

Doug Smith, KF6DX
kf6dx@arrl.org

Models and Measurements: Mandate for Change

Three years ago, a small group of engineers and scientists who are also avid hams united to improve the state of transceiver testing in Amateur Radio. Articles you contributed here and elsewhere formed some of the background for their subsequent discussions. The group identified significant discrepancies in published test data and methods, including those related to third-order intercept point, composite noise and blocking dynamic range. Also identified were several areas where tests were notably lacking or missing. The group has begun to document its findings and suggestions.[1,2,3] That series shall continue here, with some of the following points in focus.

For receivers, chief among the enemies of performance from within are noise and "birdies." We have articles coming that will deal specifically with those issues, and more.

The main issue with receiver noise testing is bandwidth dependence. Because a receiver's noise floor power sets one end of dynamic-range measurements, it would be nice if it were expressed in bandwidth-independent terms. *The ARRL Handbook* mentions normalizing noise to a 1-Hz bandwidth, as is done for phase-noise data. Alternatively, a noise figure in dB could be stated. Measurements that don't measure noise figure directly can be converted to that basis when the noise power and bandwidth of the measurement are accurately known.

Birdies are signals appearing in or near a receiver's passband that are caused by the fundamentals or harmonics of internal oscillators or the unwanted mixing products of them. They're largely ignored in formal receiver testing now, but they're as important to potential buyers as other parameters. They need to be predicted and examined.

Other enemies from within that need more regular attention are things like frequency accuracy and stability, leakage of local oscillators and other signals to the antenna, automatic level control and its effect on transient occupied bandwidth, Part 15 and CE compliance, and reliability, serviceability and warranty of equipment. All those things are of interest to buyers and users alike.

Enemies to performance from without depend on conditions outside the unit under test. The bandwidth dependence of receiver dynamic-range testing should be removed. Every test must measure what it's supposed to measure. Receiver spurious responses other than IMD2 and IMD3 are not gener-

ally searched but they need to be explored. That's especially important for rigs using direct digital synthesis or digital direct conversion schemes. Receivers with analog-to-digital converters close to the antenna don't always behave according to traditional models. Yes, they exhibit second-, third- and higher-order distortions, but not as traditional units do. It remains to be seen what other software-radio-related limitations must be defined and measured. That will be the subject of future articles here.

Some performance measurements over ranges of supply voltages, temperature and vibration should be standard. Prospective buyers and users need to know what protection circuits a rig employs: VSWR power cutback, antenna tuner range and efficiency, reverse polarity and overvoltage protection at the power input, lightning and static discharge protection, and so forth.

Then there are certain ergonomic issues for which it's difficult to specify and design tests; but in this microprocessor-controlled age, it *is* possible to test control systems. One test is the famous "dead-man test," wherein you engage all the controls on a rig in random combinations to try to make the unit crash. The processed audio quality of transmitters and receivers, especially when special digital techniques like noise reduction are being used, is difficult to report objectively but we do have methods of converting subjective evaluations.

Measurements must be grounded in established scientific principles. They must be traceable to standards of known precision. Uncertainties must be declared for all measurements. When systematic measurement errors are discovered, previously published data must be corrected.

In short, changes and additions to existing test methods, along with corrections to already-published data, are mandated and as part of our stated mission at *QEX*, we'll be here to document them. The ARRL, for one, is taking steps in the right directions. See recent *QST* Product Review columns for details

— *73, Doug Smith, KF6DX,* **kf6dx@arrl.org**

Notes
[1]U. Rohde, N1UL, "Receiver Measurements, How to Evaluate Receivers," Jul/Aug 2005 *QEX*, pp 3–11.
[2]D. Smith, KF6DX, "Tech Notes: Quantifying Measurement Uncertainty," Jan/ Feb 2006 *QEX*, pp 54–56.
[3]L. Åsbrink, SM5BSZ, "Blocking Dynamic Range in Receivers," Mar/Apr 2006 *QEX*, pp 35–39.

QEX

I-V Curve Tracing With A PC

Design of an Inexpensive PC Sound Card Curve Tracer

Dr George R. Steber, WB9LVI

Curve tracing is an interesting way to learn more about electronic components. By curve tracing, we mean making a graph of current versus voltage (an I-V curve) to display the basic characteristic of the electronic device. The I-V curves of diodes, LEDs, transistors and other devices are very helpful in understanding their nonlinear operation. Such curves are useful in establishing circuit models as well as determining bias points, load lines and limitations of the devices. For many, curve tracing seems to be a lost art and many experimenters rely only on curves supplied by the device manufacturer.

Unfortunately, the manufacturers' I-V curves are not always handy or the devices may not be documented at all. Of course, there are commercial curve tracers that will do the job but they are outside the realm of most experimenters. In this article I will attempt to remedy that situation and describe for you an economical way of obtaining the I-V curves of two and three terminal devices using a PC, a simple circuit, a *Windows*-compatible sound card and a bit of software.

To my knowledge, no one has previously used the sound card inputs of the PC for curve tracing without resorting to fancy modulation schemes. There is good reason for this as most sound card inputs are ac coupled. That is, there is a capacitor in series with each input that effectively removes the dc value of the signal. When tracing a nonlinear device, like a diode, the actual dc value must be used or accuracy will suffer. While the PC sound card has a nice pair of 16 bit analog-to-digital (A/D) converters of relatively good accuracy they would be of little value unless this coupling problem was solved. A novel and completely new method that retains the dc value when using a sound card was developed by the author and disclosed in a recent article.[1] One outstanding feature of the procedure is that it can be

[1]Notes appear on page 9.

9957 N River Road
Mequon, WI 53092
steber@execpc.com
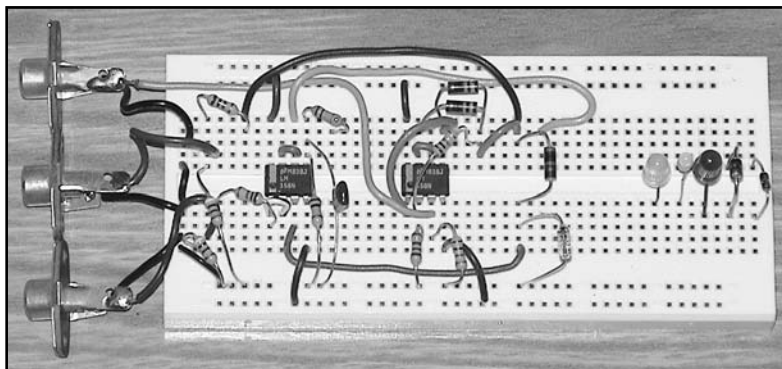
implemented without modifying the sound card or even opening up the PC case.

For this project I have designed a new circuit that is much simpler than the one described previously (see Note 1) and have also written new software to make it easier to use. An abbreviated article describing construction and operation of this new curve tracer appears in the July 2006 issue of *QST*.[2] Presented here are the technical details behind this unusual system, dubbed Curve Tracer II. In case you don't have the *QST* article handy I will also present some material on installation and operation of the curve tracer as well as theoretical details and additional application material for tracing transistors and integrated circuits. It's a project that should appeal to almost anyone involved in electronics.

This project requires a PC running the *Windows* operating system. You can use one of the newer 3 GHz PCs or dust off your old 200 MHz PC. Of course, there are no guarantees that this project will work with your system. But it has been tested with a 200 MHz Pentium Pro, a 500 MHz Pentium III, and a 1.1 GHz AMD Athlon processor running *Windows 98SE* and *Windows XP* with *Sound Blaster (SB) Live!*, *Value Edition*.

So, if you have a Pentium or AMD PC with a *Windows*-compatible sound card, you probably have the basis for a very good *Windows*-driven curve tracer. All you need to do is build the simple circuit described, connect it to your computer sound card and run the program. You will be able to accurately plot the I-V characteristics of

many two and three terminal devices such as resistors, diodes, Zener diodes, LEDs, transistors, integrated circuits and others. Data is captured via the sound card stereo input, processed and plotted on the screen. The circuitry is very low cost (less than $2.00) and uses readily available parts. You can build the circuit on a solderless breadboard like I did, or design a circuit board for it. In any case you will need a digital voltmeter (DVM) for calibration purposes. To get started it will be helpful to review some concepts of curve tracing.

DC and AC Curve Tracing

Curve tracing with dc signal levels is simple in principle, and requires an adjustable dc power supply, ammeter, voltmeter and current limiting resistor, R, as shown in Figure 1A. In this case only a positive voltage is shown, but both polarities are often used for tracing. One adjusts the supply voltage while monitoring the meters and writes down the meter readings as a table of I and V values. From the table, data points can be plotted to obtain a graph, like the one shown at Figure 1B. While this works well, it has several drawbacks. One has to manually adjust the voltage at reasonable intervals, read and write down the data, and finally plot it. You also have to switch power leads if you want to plot negative voltages. If more than a few devices need characterization, an automatic method is preferable. Clearly, if you have a bag of diodes, LEDs or Zeners that you just bought at a hamfest and want to test them, this is not the way to do it.

Classic ac curve tracers are faster, but slightly more complicated. You apply an ac voltage to the device under test (DUT) and plot its current versus voltage on an oscilloscope. You need an ac voltage source, transformer, a few resistors and, of course, the oscilloscope. A straightforward curve tracer circuit is shown in Figure 2A.

Transformer T1 is used to provide isolation and voltage reduction from the ac source, usually 120 V ac, 60 Hz. Resistor R_1 limits the current while R_2 is used to measure current. An oscilloscope is used in dc-coupled x-y mode for display. The scope x input measures the DUT voltage and the scope y input measures the voltage across R_2, which is proportional to current. For example, if R_2 equals 100 Ω and there is 1 V across it, it follows that there is 10 mA through the DUT. This means the y channel is calibrated for 10 mA per V. Notice that since the ground point for the scope is at the connection of R_2 and the DUT, the current signal measured by R_2 will be inverted. This can usually be corrected at the oscilloscope by reversing the polarity of the channel. Notice that the ac tracer sweeps both positive and negative voltages so a diode is sometimes placed across the DUT to provide reverse voltage protection. A typical I-V curve is shown at Figure 2B.

While both curve tracing approaches are straightforward, there are some things to consider. The dc curve tracer is painfully slow and prone to errors. The ac curve tracer is faster but requires a transformer and an oscilloscope, which requires specialized knowledge to use and calibrate properly. Since we are often saving, comparing and printing I-V curves, a PC can be of considerable help. So, now, let's consider how we can use a PC and a sound card for curve tracing.

Sound Card Software Model

Modern PC sound cards are marvelous examples of technology. They are low cost, low noise and feature high sampling rates, operating well into the high audio range. Most impressive is that they have dual 16 bit A/D converters, dual 16 bit D/A converters and can operate at full duplex. While there are a few cards that are dc coupled, the vast majority of them are ac coupled. This means they can't be used directly in many data acquisition tasks requiring measuring dc voltages. Hence, most people have given up using them for this type of application or have conjured up fancy modulation schemes that require ac carriers. This article describes a method that measures dc voltages with your ac sound card, without modifying it. This method uses a software model of the ac input circuit to predict and reverse the effects of the ac coupling. Here is how it works.

AC coupling is accomplished by means of large capacitors, which allow the frequency response to go down to very low frequencies, typically 20 Hz. A generic model of one channel is shown in Figure 3A. Figure 3B shows the step voltage response. The signal rises quickly to the value of the step, then decays exponentially to zero. If you apply a step voltage to your sound card input, record it, and view it, that is what you will see. The voltage transfer function of this circuit, using the Laplace variable, s, is given by:

$$\frac{V_2}{V_1} = G_1(s) = \frac{R_1}{R_1 C} \qquad \text{[Eq 1]}$$

Now consider the addition of another transfer function, $G_2(s)$ in cascade with $G_1(s)$ as shown in Figure 3C. If we could make $G_2(s)$ equal to $1/G_1(s)$, the overall transfer function would be unity and we would see a flat, non-decaying step response at point V_3. Notice from Equation 1 that there is a zero at the origin. This forces $G_2(s)$ to have a pole at the origin, making it marginally stable. If we limit our time interval, however, the potential instability can be managed quite well.

Since we don't want to open the PC and add the circuit for $G_2(s)$ to the sound card, what can we do? We can build a software model of $G_2(s)$ in the computer. Basically we
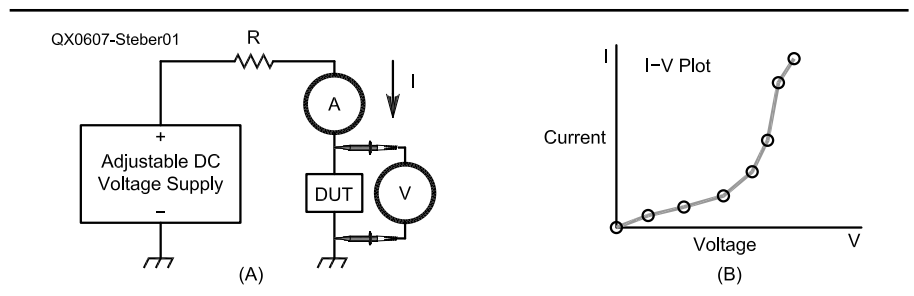

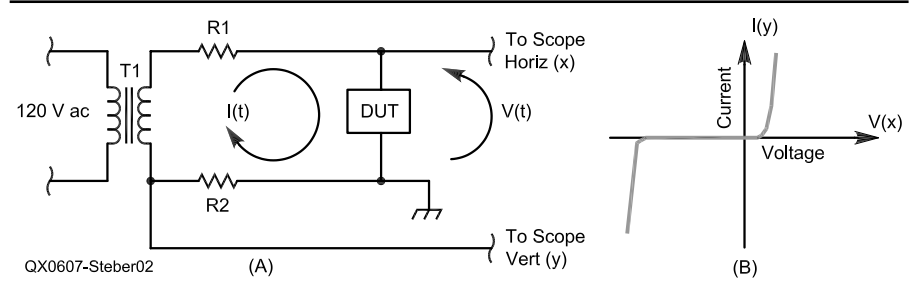Figure 1 — Simple dc curve tracer setup and I-V plot example.

Figure 2 — This schematic shows a basic ac curve tracer setup.
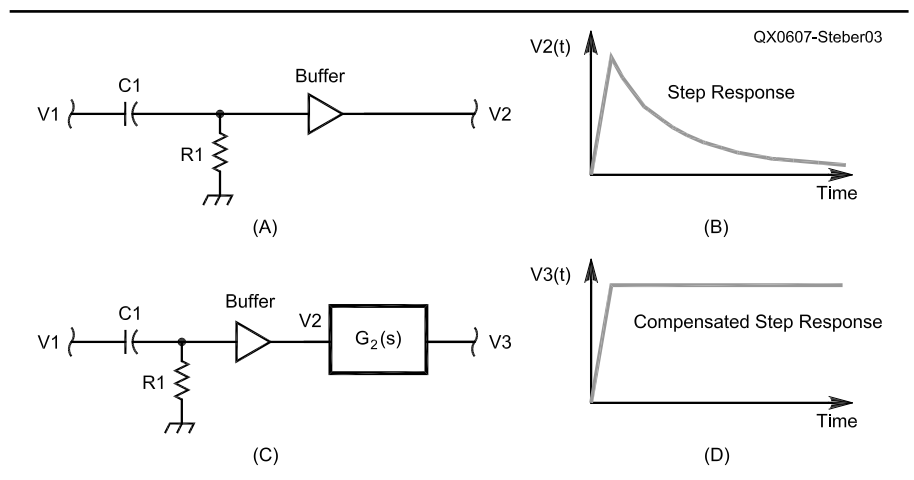

Figure 3 — The circuit at A shows a typical sound card input and B shows the step voltage response. The circuit at C shows the addition of a hypothetical compensation circuit and D shows the corrected step response.

use a discrete-time digital filter in software that recursively solves a difference equation given the initial conditions and input data sequence from the A/D. In effect, we use a sampled version of $G_2(s)$, which will now be called $G_2(z)$. How can we design a digital filter that satisfies the specifications of the analog filter $G_2(s)$? Mathematically we use the following bilinear transformation:

$$s = \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \qquad \text{[Eq 2]}$$

where T is the sampling period and z is the z transform of s. In our case, the resulting $G_2(z)$ is given by the following:

$$G_2(z) = 1 + \frac{T}{2R_1C_1} \frac{1 + z^{-1}}{1 - z^{-1}} = \frac{V_3(z)}{V_2(z)} \qquad \text{[Eq 3]}$$

To prevent aliasing, we must limit the input frequencies to less than ½ of the sampling frequency. Implementing the digital filter given by Equation 3 on the sound card data will correct it and compensate for the effect of the coupling capacitor. One slight problem remains; we don't know the value of the product R_1C_1. We can determine this parameter of the filter experimentally, however. If we apply a step input to the sound card, the resulting output can be observed on the screen. We can then adjust the parameter of the filter until we get a flat step response. Knowing the exact value of the step amplitude, as measured with our DVM, also allows us to accurately calibrate the sound card dc voltage.

This procedure should work well if we limit the time duration of the input. In practice it works well for a curve tracer.

Sound Card Interface Circuit

The simple ac curve tracing circuit mentioned earlier is not appropriate for this approach since it doesn't provide a step output for calibration and it can't be started and stopped. So a special interface circuit was designed for this project. That circuit is shown in Figure 4.

A few general comments are in order. Sound cards typically have a 3.5 mm stereo jack for input and output so you will need a cable, preferably shielded, to connect between the card and the circuit. Actually, you will need two stereo cables (one for input signals and one for output signals) with the 3.5 mm stereo plug on one end and the appropriate plugs (like RCA plugs) on the other end. Such cables are commonly found at your local electronics or AV store.

Regarding the circuit, make sure to use LM358 op amps. Substitutes like the TL082 or TL272 will not work well here. Also, use a good, well-regulated dual power supply of ± 12 V with at least 100 mA capability. The DUT may require a pulsed current of 60 mA or more, and the supply voltage should not droop during this time.

The interface uses one sound card output (either left or right channel) and two sound card inputs (the left and right channels). Although the schematic shows the inputs separately for clarity, they are in fact combined into a single stereo jack. The line output will provide a special sine wave as described later. Resistor R_1 is there to provide a ground reference for the sound card line output. U1A and U1B form a dc power amplifier to boost the line output and provide more current capability. This power amp does not provide equal current source and sink capability since the LM358 is unbalanced in this respect. Typically an LM358 op amp can source 40 mA and sink 20 mA. Since we are using two op amps in our power amplifier, we should be able to double those figures. My breadboard circuit actually produces less, sourcing only up to 65 mA and sinking up to 32 mA. Nonetheless, this is a stable, overload-protected circuit with little dc offset, which is why I used it. Observe that in the case of diodes or LEDs, one can always orient the DUT so that current is sourced to the DUT, if desired. Note that resistor R_2 controls the gain of the power amp. If you need to increase the gain, decrease the value of R_2.

U2B is a simple step generator, providing a slow step voltage, at about ½ the positive supply voltage, at test point TP1 when S1 closes. Use a single-action *momentary* switch for S1 and avoid using a slide switch. Jumper block J1 is used to select the step or sine wave signal for the sound card. The step voltage is only used once to calibrate the system. Note that there is also a test point, TP2, near pin 1 of J1. With no signal applied to the power amp, the voltage at TP2 should be less than 5 mV dc. This will avoid having a dc voltage applied to the sound card inputs
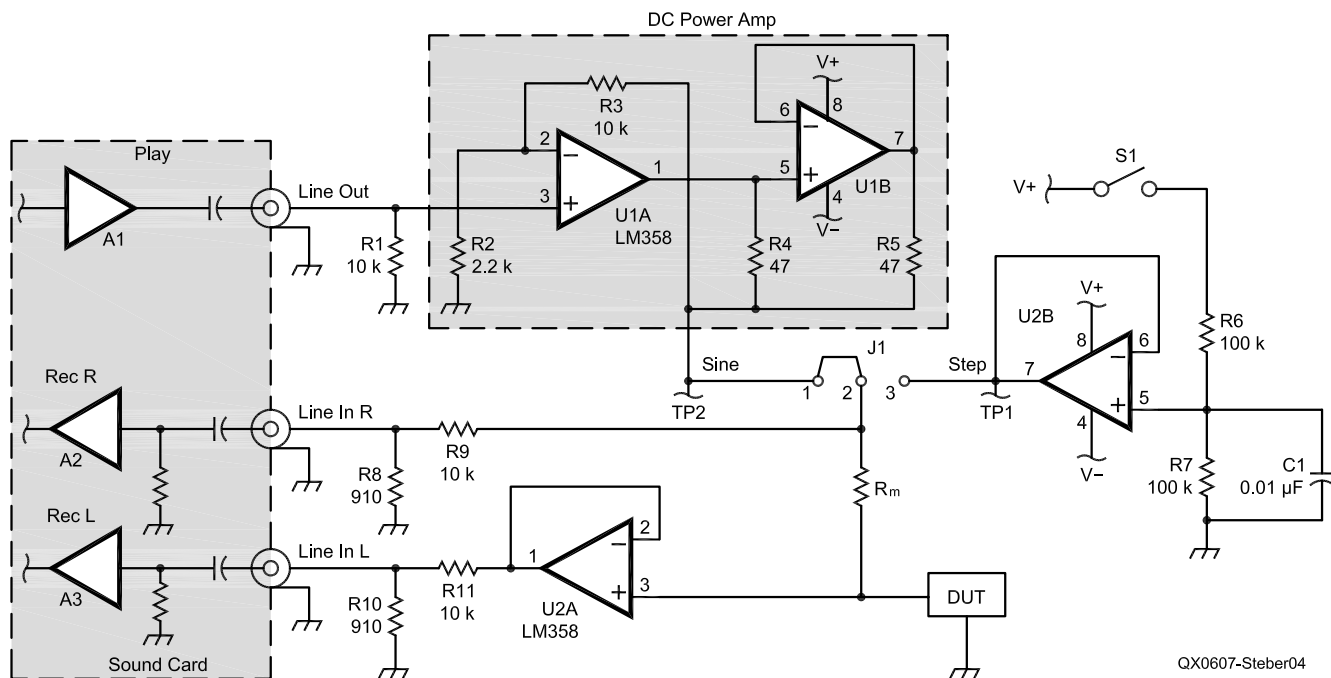

Figure 4 — Curve tracer circuit and the interface to the PC sound card.

and causing an error. In my circuit, the voltage at TP2 is around 3 mV.

A current measuring resistor, Rm, is connected in series with the DUT. This resistor is typically 100 to 1000 Ω. The exact value can be chosen to suit requirements. R_8 and R_9 form an attenuator to reduce the voltage to something the sound card line input *right* channel can accept without overloading, typically around 1 V. (My *SB Live* begins clipping at 0.82 V.) The voltage across the DUT is buffered by U2A, similarly attenuated and fed to the *left* channel. The *current* through the DUT is the difference between the right and left channel voltages, divided by Rm. So we need to know the value of Rm accurately, say within 1%. Given the value of Rm, the software calculates the current in the DUT.

The circuit is potentially harmful to some devices, since in normal operation the sine wave sweeps through both positive and negative voltages up to 10 V. To restrict the voltage sweep to only positive or negative excursions, a diode can be appropriately placed across the DUT. The sine wave may flattop as more current is drawn from the power amplifier by the DUT. This is normally not a problem. Increasing Rm may help reduce this effect. This circuit can easily be modified to handle higher voltages and currents. It is only a starting point. As it stands, a large number of devices can be curve traced with this circuit.

PC Curve Tracer Operation

Basically the curve tracer works by applying a test signal, a few cycles of a 240 Hz sine wave, to the DUT and measuring the current through the device and the voltage across it. Referring to Figure 4, the shortened sine wave is generated in the computer and is output to the line output of the sound card amplifier A1 when the Start button on the computer screen is clicked. The sine wave is boosted by the dc power amplifier and presented to the top of Rm via pin 2 of J1. This sine wave produces a current in the DUT and a voltage across it, just as in the simple dc curve tracer discussed earlier. The voltage at the top of Rm is sent to the right channel of the sound card and the voltage at the bottom of Rm is sent to the left channel. Knowing the voltage across Rm enables the computer to calculate the current using Ohm's Law. Notice that the voltage at the bottom of Rm is also the voltage across the DUT, which is sent to the left sound card channel.

As mentioned earlier, we need to calibrate the software model so that it exactly reverses the effects of the ac coupling. For this purpose we use a specific test input, namely a step voltage and the circuitry associated with U2B. With the DUT removed and the jumper on J1 set between pins 2 and 3, we start a software

capture. Shortly after clicking Start we initiate a step voltage by pressing S1. Holding S1 closed, the step voltage amplitude is also measured with a digital voltmeter at TP1. The program captures the step voltage and is now ready for calibration of the software model, given some input from you. More specific details on how to do this are given later. The calibration is then saved and used for all further curve tracer operations. The calibration values have been found to hold up well over many months without drift, but can always be redone when needed.

Sound Card Considerations

A low-distortion, low-noise, full-duplex sound card is desirable. The economical *Sound Blaster Live, Value Edition* fills the bill nicely and probably many others will too. Since I cannot test them all, I will restrict my attention to this one. Referring to Figure 4, we see that A1 is the line output amplifier, and A2 and A3 are the right and left channel line inputs of the sound card. The mixer control in your *Windows* software controls the gain of these amps. The level of the line output signal is controlled in the Play section of the mixer via the Wave and Spkr sliders. So to control the amplitude of the sine wave output, just adjust the sliders.

The gain of the line input amplifiers A2 and A3 are controlled in the Record section of the mixer by the Line slider. They will saturate if the input voltage is too high, regardless of the Record setting in the mixer. On the SB, this occurs at 820 mV. Fortunately, this is easy to detect, because the captured signal will flattop. Since the linear gain of the amplifier is lost at saturation, we may also see a fictitious current bump in the DUT current trace at the extreme voltage limits. A real time oscilloscope function is provided in the software to help monitor for this

condition. So, if you see a current bump in the I-V trace when there is no DUT in place, it is likely caused by an amplifier saturating. Just back off a little with the Play output signal or just ignore it.

A note is in order on earlier Sound Blaster sound cards such as the SB16 and AWE 32 since there are so many of these still around. Unfortunately, they do not provide true full duplex operation and are noisy. The same may be said of SB *compatible* cards, so be wary. For example (with the latest drivers) the SB AWE32 can only play unsigned 8 bits and record signed 16 bits at once. It also has a built-in amplifier that may overdrive the interface circuit. My advice is not to use any of these cards.

Curve Tracer Software Installation and Operation

The Curve Tracer software is available on the *QEX* Web site and is zipped for fast downloading.[3] Unzip it to a new folder and you are ready to go. Just run the executable (exe) program. It was tested with *Windows 98* and *XP*. When you run the software you may get a message like "Required DLL file MSVBVM60.DLL was not found." This is a Visual Basic run time file and is on many systems. If not found, you will need to obtain it and install it on your system. It is freely available from Microsoft and other sites on the Web. It is usually available as Visual Basic 6.0 SP5: Run-Time Redistribution Pack (VBRun60sp5 .exe) and is a self-extracting file. Download takes about 6 minutes at 28.8 kbps.

If you get the message "Component 'COMDLG32.OCX' or one of its dependencies is not correctly registered: a file is missing or invalid." when you try to run the program, you will need to register it on your system. This file is also available free from Microsoft and other sites on the Web. More
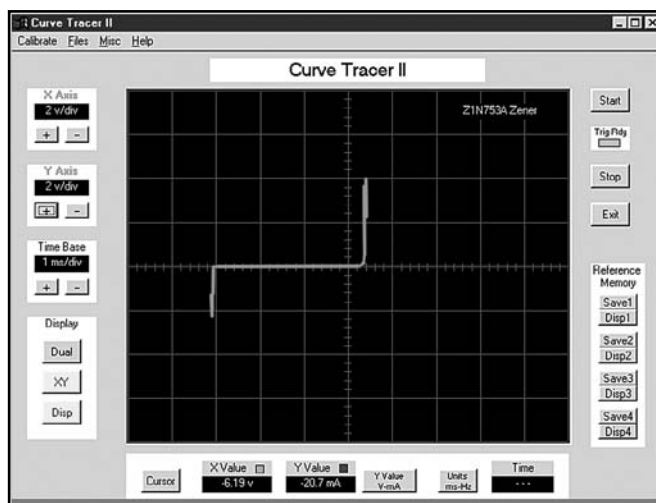

Figure 5 — The main window of the curve tracer program showing Zener diode plot.

details are included with the software.

If you just want to experiment with the Curve Tracer program, go ahead, as it does not modify the registry or install any other material on your computer. You can remove it by just deleting the entire folder where you copied the software.

Figure 5 shows a screen shot of the Curve Tracer II. The oscilloscope display is in the middle of the screen. It functions like a real oscilloscope and has controls for dual trace and X-Y display as well as sensitivity settings on the left side. The X-Y mode is essentially the I-V mode with current on the vertical axis and voltage on the horizontal. The *origin* is at the center of the screen. This screen shot shows the I-V display of a 1N753A Zener diode. It is rated at 6.2 V at 20 mA. by the manufacturer. I measured 6.19 V at 20.7 mA with my Zener.

Below the scope screen is the cursor readout, which functions with the on-screen cursor and is controlled by your left mouse button. It works in both dual and X-Y mode. On the right side are the Start button, Stop button and TrigRdy LED. The Start button initiates a capture, which automatically stops after 1 to 2 seconds while processing the data and displaying it on the screen. In case something is wrong, like you forgot to turn on the circuit power supply, the Stop button may be used to stop the capture. Below the Stop button are four temporary memories for saving and comparing data. Additional features include a second cursor (right mouse button) for time measurement, provisions for printer output, screen capture and file saving for use by *Excel* or other programs.

System Setup and Calibration

First make sure the circuit is powered and connected to the sound card line inputs, as shown in Figure 4. Check the mixer program that came with your sound card or the one that came with *Windows* because you may have changed its settings. When you start the program, a little notice comes on the screen to remind you to do this. Basically you want to set the output level, input gain and stereo balance. The details on how to do this will obviously vary from system to system. Here is how it's done with the SB.

In the mixer Play section, enable Wave and Spkr, set the sliders to maximum and mute all others (including Line to avoid audio feedback). In the Record section, enable Line, set it to maximum and mute all others. Set the stereo balance to center for all controls. These settings are *important*. For example, if you forget to mute Line in the Play section, there will be errors.

Calibration needs to be performed only once since parameters are saved in a file CTinit upon exit. Follow these steps closely so good accuracy can be achieved. First,

remove the DUT. Next, set the jumper J1 on the circuit so that the step voltage is applied to Rm. Run the program. Click Start and wait for the TrigRdy LED to change from green to red. Now create the step voltage by closing switch S1 on the circuit. Upon capture of the step voltage data, the TrigRdy LED will revert to green. With S1 still closed, measure the step voltage at TP1 with a good voltmeter, preferably a digital type, and write it down. View the data captured by the program using the left cursor in *dual* mode. You should see a large step voltage on the left (L) channel (green) and a small voltage on the right minus left (R – L) channel (blue). See Figure 6 for an example. [The blue line on the computer screen is shown as a heavy horizontal line on Figure 6. — *Ed.*] The data should start out clean without much noise. If there is a lot of noise at the start of capture (probably due to S1 bounce), repeat the capture until it is clean.

The captured data is used to calibrate the program. Click on the Calibrate button to begin. Change the values shown in the Calibrate box as needed. All calibration is done on the originally captured step voltage. It does not need to be recaptured each time. Use the left cursor (left mouse button) to measure the data on the L and (R – L) channels as needed.

Adjust the Channel Balance value until the (R – L) channel is close to zero. This is essentially the voltage across Rm, which should be zero since the current in it is zero because the DUT is not present.

Adjust the Compensation value until there is no droop or rise of the step voltage of the L channel. After its initial rise, the step should flatten out to be a straight horizontal line across the window of the scope. If it is over corrected, the line will rise while if under corrected, it will decline.

See Figure 6 for an example of over, under and normal compensation. Make corrections in *very* small amounts. Use the cursor to verify it is straight by moving it across the waveform while looking at the readout.

The volts-per-bit calibration is done on the L channel of the sound card. Use the left cursor and measure the voltage of the L channel. Adjust this value until the cursor value is close to the value you measured with the voltmeter earlier. Each time you click the Apply button in the Calibration box, these values are applied to the captured step data. You will immediately see the results on the screen. Repeat these steps until the best calibration is achieved.

Other Calibration Box Values

The Trigger Level controls the voltage level at which the beginning of the input signal is captured. If you get false captures, you can increase this value. If the value is too high, you will miss the leading edge of the captured waveform. If it is too low, you will get false triggers due to noise.

The reason for the Invert Input box is that some sound cards invert the signal. The SB Live inverts the data, so this box needs to be checked to correct the input polarity. To determine if your sound card inverts the data, look at the step waveform of the L channel. It should start out positive for a positive input such as provided by the hardware circuit. If not, you need to check the box to invert the data.

Enter the value of the current measuring resistor in ohms in the Rm box. The cursor readout will display device current based on this value using Ohm's Law.

Here are values for my SB Live card. Channel Bal: 0.9551; Compensation: +10; Volts Per Bit: 0.00033982; Trigger Level: 200; Rm: 98.8 and Invert: Box checked.
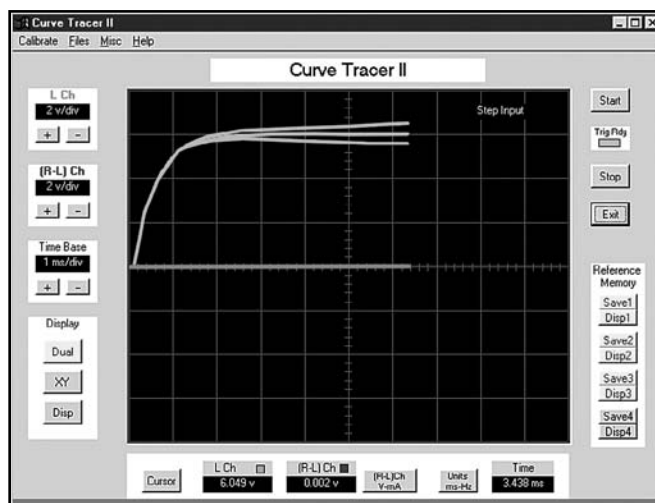

Figure 6 — The main window of the curve tracer program showing step voltage input.
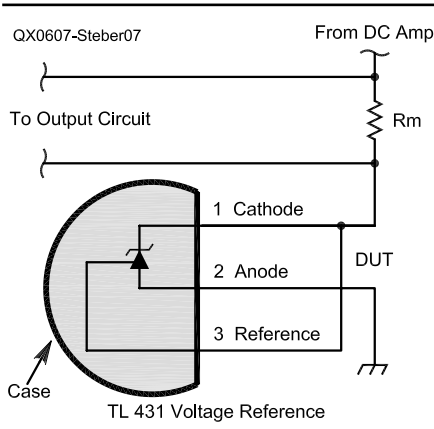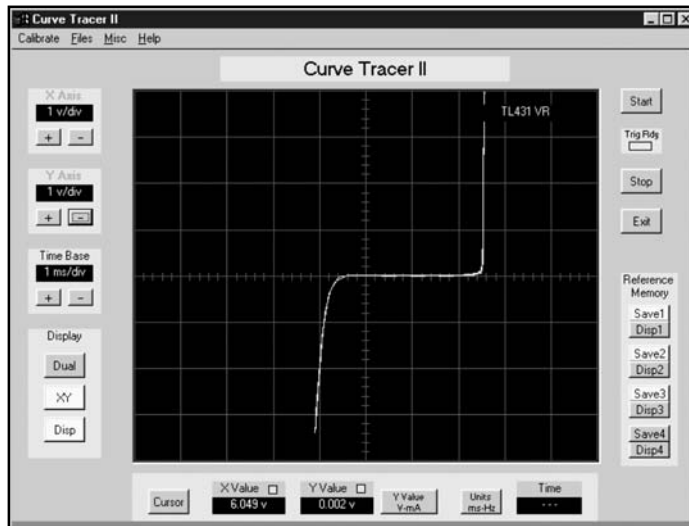
Figure 7 — Voltage reference diode circuit.

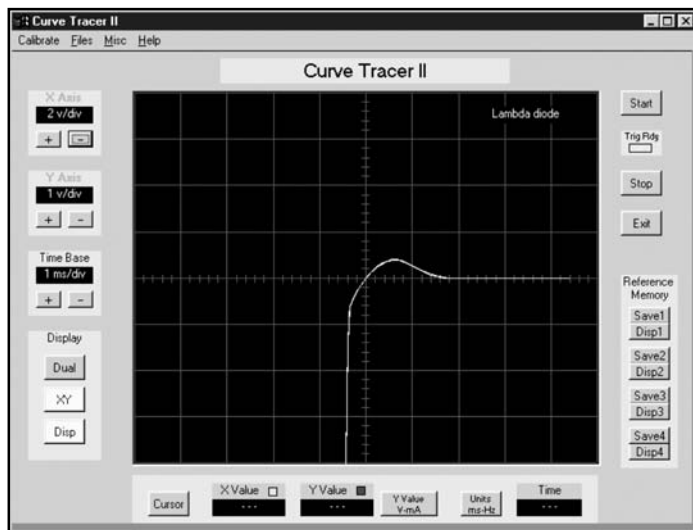Figure 8 — Voltage reference diode I-V curve.

Figure 9 — Lambda diode I-V trace.

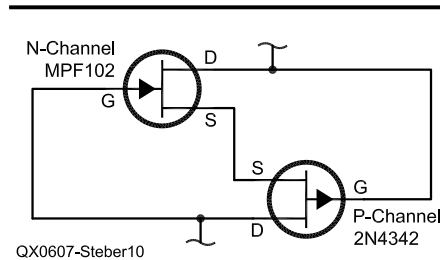Figure 10 — Lambda diode circuit.

of the TL431 is shown in Figure 8.

Speaking of diodes, a strange device called a lambda diode makes an interesting I-V curve as shown in Figure 9. Notice the large negative resistance region to the right of the origin. This device has been found to be useful in oscillator circuits. You can make your own device, like I did, by combining back-to-back N-channel and P-channel FETs as shown in Figure 10.

Curve Tracing of Transistors and Integrated Circuits

Figure 11 shows the collector curves of a 2N3904 transistor. Multiple curves are displayed using the reference memory, with each curve saved separately. You need to provide a simple base current source for transistors. Such a base current circuit is shown in Figure 12. R1 is chosen to be very large to essentially form a current source to the base. Adjustment of R2 varies the base current. Normally you would not trace the negative voltage region as done here as it might damage the device. In this case it is interesting to see the negative resistance region of a reverse biased transistor (on the left side of the origin) and fortunately the device was not damaged. FETs can be traced in a similar manner.

Obtaining I-V Traces

It is easy to curve trace a component. Make sure the hardware circuit is connected to the sound card, J1 connected to *sine*, the program calibrated, and the DUT connected. So you know what to expect, use a known resistor, diode or Zener as your first DUT. To curve trace it, follow these steps.

Click the Start button. The green TrigRdy LED will *not* turn red as the sine wave generated is only a few cycles long. After 1 to 2 seconds the data will be captured, processed and will appear on screen. You have now curve traced your first component! If for some reason, the trigger did not occur, the TrigRdy LED will turn red and can be manually reset using the Stop button. Normally this is not required unless there is something wrong with your setup. You cannot exit the program unless the TrigRdy LED is green, a precaution to make sure all processes are shut down.

Curve tracing various components is interesting and instructive. Resistors, of course, plot as straight lines. Their value can be checked by choosing a point on the line with the cursor (yielding the voltage and current at a given point) and using Ohm's Law. Diodes, Zeners, and LEDs show plots similar to Figure 5. Capacitors and inductors do not plot as their characteristic elliptical shape because there are not enough cycles in the test sine wave to reach steady state.

An interesting practical application is checking of voltage reference devices like the TL431, LM336Z and others. Figure 7 shows a circuit for testing a TL431 programmable reference device. The typical reference voltage given by manufacturer is 2.495 V at 10 mA. My device measured 2.52 V at 11 mA, which is within tolerance. Such reference diodes are also useful for checking the veracity of the curve tracer. The I-V curve

Figure 11 — Collector I-V curves of 2N3904 transistor.

Figure 13 — Integrated circuit test setup. [Normally any unused inputs would be tied to Vcc. In this example, the device was tested with no connection to the second input or to the output —*Ed.*]

QX0607-Steber13

Figure 12 — Simple base current circuit for transistor.

Figure 14 — Sample I-V curve of 74AHCT00 integrated circuit.

Final Thoughts

Files of several components that I have traced and other curves are included in the zipped file package mentioned earlier. You can use the program to read them and see what a trace looks like without having to build the circuit. There are probably many more applications for this curve tracer, perhaps like analog signature analysis. Let me know if you come up with an interesting application.

The Curve Tracer II was compared to an ac curve tracer built using my Tektronix TDS360 Digital Real Time Oscilloscope and the results were in good agreement. With any luck, you may achieve similar results. In any case, I hope I haven't thrown too many *curves* your way!

Notes
[1]George Steber, "Tracing Voltage And Current," *Circuit Cellar*, January 2004
[2]George Steber, "A Low Cost Automatic Curve Tracer," *QST*, July 2006, pp 32-36.
[3]You can download the software associated with this article from the ARRL Web at **www.arrl.org/qexfiles/**. Look for 7x06Steber.zip.
[4]Tom Mathews and Timothy Toroni, *Design Feature*, "Rework Within Your Reach," September 2, 2004, *EDN*, pp 75-81.

George R. Steber, PhD, is Emeritus Professor of Electrical Engineering and Computer Science at the University of Wisconsin-Milwaukee. George, WB9LVI, has an Advanced class license and is a Life Member of ARRL and the IEEE. Dr Steber has written several articles for QST, *the most recent one being "A Low Cost Automatic Impedance Bridge" in the October 2005 issue and a companion article in the Sep/Oct issue of* QEX, *"An LMS Impedance Bridge." He has industrial experience as a company officer, consultant, and product designer, with 18 patents issued. In his spare time he enjoys racquetball, reading, editing video, astronomy and playing classical, Broadway and big band favorites with his trumpet in a local band. You may reach him at* **steber@execpc.com** *with "Curve Tracer" in the subject line. Send your message in text mode.*
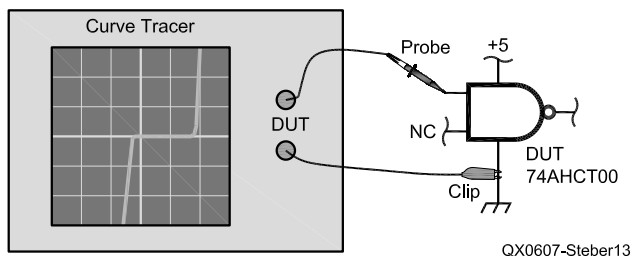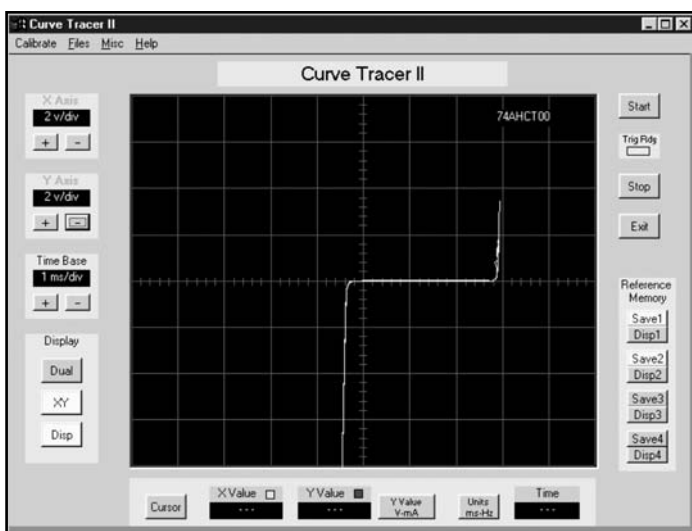
Curve tracing of integrated circuits under various conditions has been described in a recent article for purposes of rework.[4] Tracing the inputs and outputs of logic ICs — under power on and off conditions — produced some unusual curves. So I decided to try it myself. Consider the 74AHCT00 IC shown in Figure 13. Curve tracing between one of the inputs and ground produced the curve shown in Figure 14. Curve tracing of integrated circuits is a new technique, at least to me. So if you endeavor to try this, be prepared for unusual curves and possible destroyed ICs. You may learn how to test and repair some circuits during the rework process, though, as described in the article of Note 4.

Transverter RF/IF Switching Using GaAs MMICs

RF switching MMICs are a low-current, inexpensive alternative to relays for VHF/UHF and microwave transverters. Requiring only microamps at a few volts, these ICs are a good choice for battery powered equipment.

Steve Kavanagh, VE3SMA

Introduction

Many modern transverter designs employ separate mixers for the transmit and receive chains. For homebrew transverters, however, it is often more convenient to use a single mixer and image filter. This allows the use of an available surplus mixer or filter, eliminates the local oscillator power divider and reduces the amount of filter tuning needed. This comes at the expense of more complex RF and IF switching, traditionally involving many relays or sometimes circulators (in the microwave bands). The result can be quite bulky and can take a lot of dc power to run the relays. A more modern alternative is to use FET switches; RF switch monolithic microwave integrated circuits (MMICs) based on GaAs FETs are now readily available that work up into the

141 Woodside Ave, Apt 314
Cambridge, ON N1S 4R8
Canada
ve3sma@rac.ca

microwave range. This article presents examples of circuits using such devices.

Background

Figure 1 shows a typical block diagram of a transverter using a single mixer and image filter, which do double duty in both receive and transmit. Appropriate attenuators and amplifiers are switched in as needed for receive or transmit operation. Attenuator 1 is in line all the time and provides protection of the transverter from accidental transmission from the IF rig with the switches in the receive position. Attenuator 2 sets the appropriate transmit level for the mixer. Amplifier 1 brings the output of the mixer up to the desired transmit power. Amplifier 2 is the receive preamp, while Amplifier 3 is a receive IF amplifier. Splitting the receive gain between RF and IF amplifiers is usually the most economical way of achieving both good noise figure and high dynamic range. Normally the dc power to the amplifier stages is switched, as well as the RF paths, to minimize overall power supply requirements.

In a very low power transverter, all four RF switching functions can be carried out using switch MMICs. The circuits that follow employ the Skyworks AS169-73 in all four locations. This device behaves very much like a single-pole double-throw relay (most manufacturers also sell DPDT versions). The only differences are that there are two control inputs, whose voltages to ground affect the state of the "contacts," and the RF pins must be isolated from ground by blocking capacitors. This is typical of most current GaAs RF switch ICs. The typical performance from the AS169-73 data sheet is summarized in Tables 1 and 2.

In comparison with a good microwave relay, the insertion loss is a little higher for the MMIC, while the isolation is substantially lower, as is the power handling capability. Thus for output powers above a watt or so it is necessary to use a regular relay (or PIN diodes) for the antenna switching. It is important to note that the compression and intercept points are normally significantly lower at lower frequencies (and often
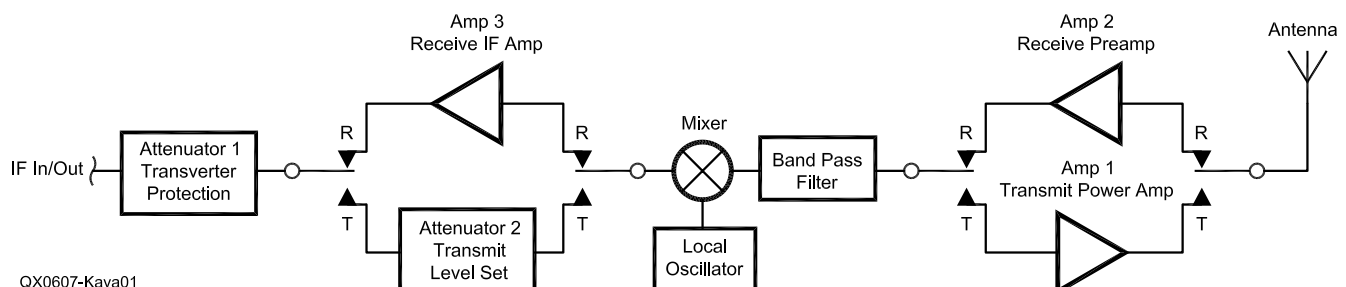

Figure 1 — Single-mixer transverter block diagram.

unspecified), so for IF switching it is necessary to keep the power levels quite low to avoid distortion. On the plus side the dc power needed to switch the device is only 1 mW (5 V × 200 μA), much lower than any relay. This can be very helpful in battery powered portable equipment.

MMICs similar to these are now quite available. A recent Digi-Key catalog lists an NEC GaAs switch MMIC, for example. Similar devices are made by many semiconductor manufacturers and often cost about the same or less than an inexpensive PC-board-mount relay, much less than a good microwave relay. In addition, if a compact layout is your goal, they are much smaller. (The AS169-73 is in a surface-mount package about 0.12 inch × 0.07 inch × 0.05 inch high, not including leads.) Be aware that some of these switches require negative control voltages, which may not be convenient. Virtually all of them come in small surface mount packages — be prepared for some fine PC-board work.

Table 1
Summary of AS169-73 Typical Performance with 5 V Control Voltages

Parameter	Typical Performance
Insertion Loss	0.3 dB (300 kHz - 1 GHz), 0.4 dB (1 - 2.5 GHz)
Isolation	25 dB (300 kHz - 1 GHz), 24 dB (1 - 2.5 GHz)
VSWR	1.15 (300 kHz - 1 GHz), 1.3 (1 - 2.5 GHz)
Input Power for 1dB Compression	+34 dBm (0.5 - 2.5 GHz) — equivalent to 2.5 W
3rd Order Intercept Point	+50 dBm (0.5 - 2.5 GHz)
Control Pin Current	200 μA maximum

Table 2
Control Truth Table for AS169-73 (pin 2 grounded)

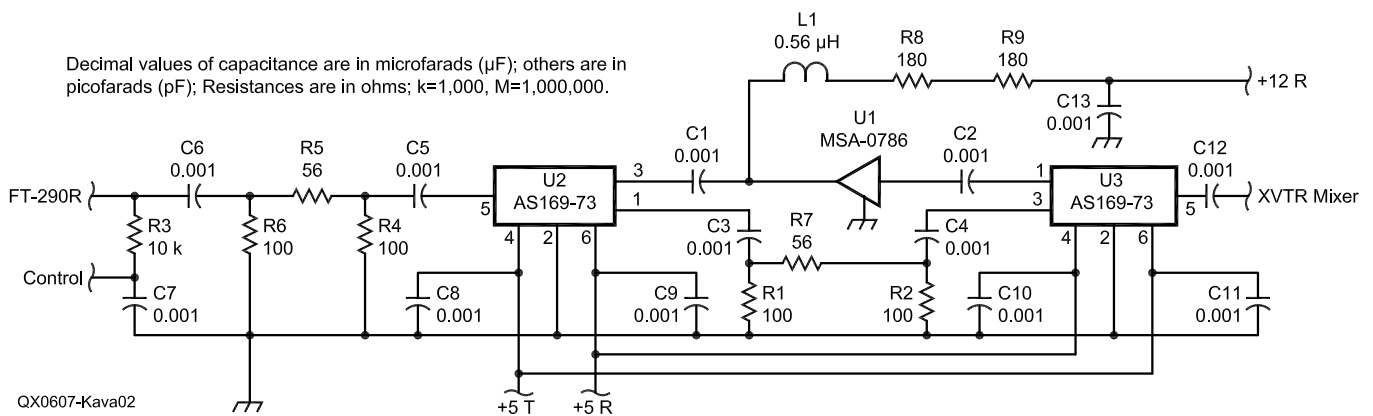| Pin 4 Voltage	Pin 6 Voltage	Pin 5 – Pin 1	Pin 5 – Pin 3
+5 V	0	Closed (Insertion Loss)	Open (Isolation)
0	+5 V	Open (Isolation)	Closed (Insertion Loss)

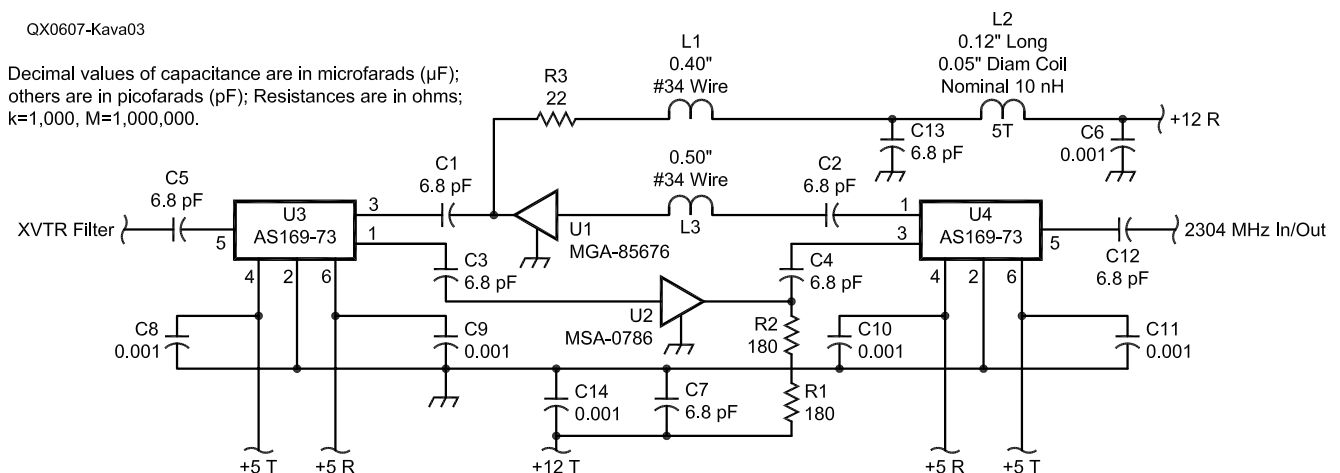Figure 2 — 146 MHz IF interface circuit.

Figure 3 — 2304 MHz amplifier board.

A Practical Example

The circuits shown in Figures 2, 3 and 4 are those I have used in an intermediate transverter, which forms part of my 24 GHz station. The IF rig is a Yaesu FT-290R (original model), which provides about 100 mW (in low power setting), as well as a few volts dc at the antenna socket, in transmit. These circuits use this dc voltage to control the transverter transmit/receive switching. The transmit output (about 1 mW) and receive input of the intermediate transverter (connected to the 24 GHz mixer) is at 2304 MHz.

Full construction details are not given since this paper is intended more to bring the attention of hams to this technology alternative. Most of you who have the skills to design and build a transverter would probably be better off working out your own approach to fit your own needs.

Figure 2 shows the IF switching circuit and Figure 3 shows the 2304 MHz switching arrangement, while the control circuit is given in Figure 4. The IF and RF circuits were made as surface mount boards (using only surface mount components), with the IF board using FR-4 material and the RF board using 0.010-inch Duroid 5880. The thin 0.010-inch board material was needed so that the 50 Ω traces used for all the RF connections (except L1 and L3) would be
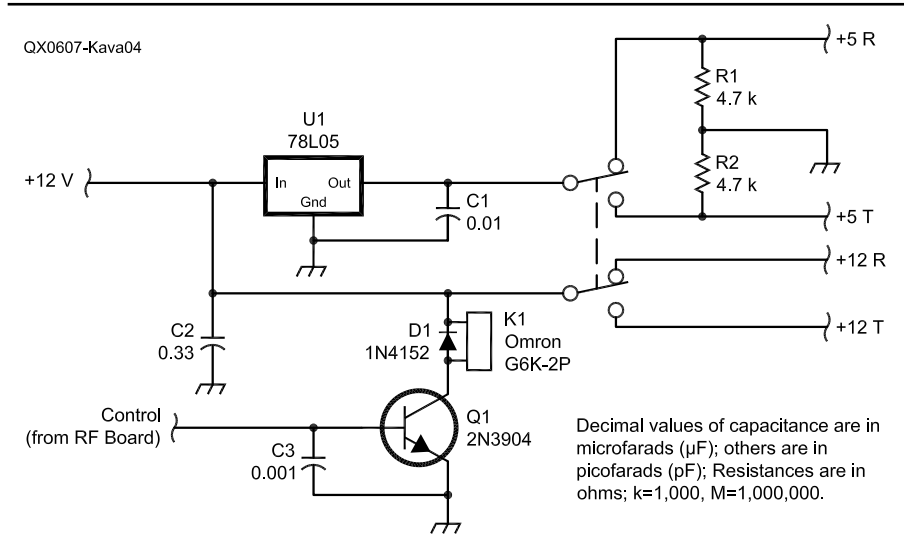

Figure 4 — Transmit/receive control circuit.

narrow enough to reach the tiny switch MMIC pins without large steps in line width. Inductors L1 and L3 are just straight pieces of fine (no. 34) wire lying on the (etched) board surface, which make transmission lines of about 100 Ω impedance. At 1.80 inches × 0.92 inch the 2304 MHz board is smaller than most microwave relays! For the ultimate in dc current reduction, a solid-state control circuit could have been used instead of relay K1.

Conclusions

Modern RF switch MMICs have worked well for me in my 24 GHz transverter's 146 and 2304 MHz sections. I have also used them for the IF switching (only) in my 5.7 GHz transverter. I was somewhat concerned that there might be potential for static damage to the switch MMICs when connecting the IF cable, but I haven't had this happen. (Yet!)

For 28 MHz IF switching it is likely that silicon CMOS analog switch devices could be used in a similar way. Even such common devices as the 74HC4066 or 74HC4053, for example, should work. With these it would probably be necessary to put multiple sections in parallel to get a decent insertion loss.

This class of MMICs offer an inexpensive, low-current, small-size choice for designers and builders of VHF/UHF/microwave transverter IF and RF switching circuits.

Steve Kavanagh, VE3SMA, was first licensed in 1974 as VE1BCZ. He has also held the call signs VE2BTW, KC6CQ and V63BD. He has a B. Sc. degree from the University of New Brunswick and an M. A. Sc. from the University of Toronto. Steve is an electrical engineer specializing in antennas and RF systems engineering. He enjoys contest operating on HF and VHF as well as microwave operating and construction of homebrew gear, and currently chairs the Microwave Band Planning Committee for the Radio Amateurs of Canada.

Octave for System Modeling

This article describes the use of the GNU Octave program to model an AM transmitter and receiver. With an understanding of these results you can model other radio systems.

Maynard Wright, W6PAP

In *Octave — Calculations for Amateurs* we introduced *Octave,* a useful mathematical software tool that can be downloaded for free under the terms of the GNU Public License.[1, 2] In *Octave for Signal Analysis*, we discussed using *Octave* to look at signals in the time and frequency domains and to transform signals from one domain to the other using the Fast Fourier Transform (FFT).[3] What else can we do with *Octave?*

[1]Notes appear on page 18.

6930 Enright Drive
Citrus Heights, CA 95621
w6pap@arrl.net

Mathematical tools like *Octave* can be used to build mathematical models of various physical and electronic systems. Let's build a model of an AM system, including the transmitter, the channel between the transmitter and the receiver, and the receiver. With renewed interest in AM transmission, it might be worthwhile to explore some of the technical aspects of this time-honored modulation scheme.

This will be a pretty simple model as shown in Figure 1. There will be no filtering in the transmitter. A "channel," or transmission path, is inferred between the transmitter and receiver, but we'll leave it empty for now and pass the signal from the transmitter to

the receiver unaltered.

We'll use the *Octave* functions *fft* and *ifft* to move back and forth between the time and frequency domains as we progress through the circuit. Sometimes we'll produce a frequency domain signal just for display to let us know what's going on at a particular point in the system.

Since the signals we'll be concerned with can be described in either the time or the frequency domain, we could work in either domain in any of these cases. When there is no reason to do otherwise, we generally pick the domain in which to work that minimizes mathematical complexity.
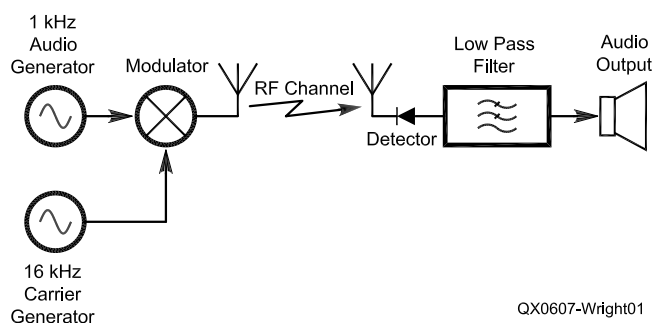

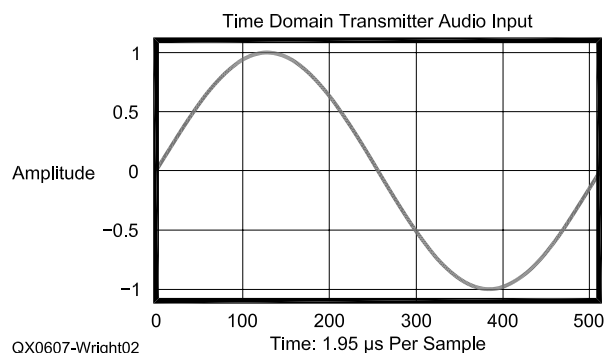Figure 1 — Block diagram of AM system model.
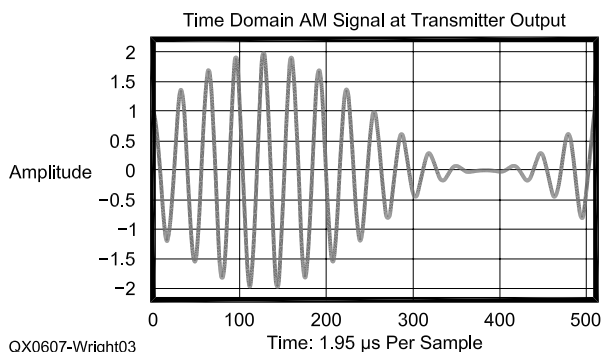

Figure 2 — 1 kHz Audio input signal.

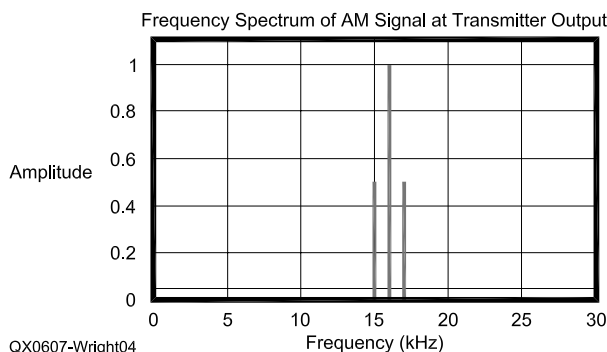Figure 3 — A 100% modulated signal, represented in the time domain.

Figure 4 — A 100% modulated signal, represented in the frequency domain.

Table 1
***Octave* Model of Plate-Modulated AM Transmitter**

```
#! /usr/bin/octave -qf

# function trunq() produces a truncated copy of the
# frequency domain representation of the signal for
# display by function plot.

function retval = trunq(spectrum)
  if(nargin != 1)
    usage("trunq(vector)");
  endif
  for n = 1:80
    retval(n) = spectrum(n);
  endfor
endfunction

# develop a vector for displaying the frequency on
# a plot

w1(1) = 0;
for n = 1:79
  w1(n + 1) = 0.5 * n;
endfor

# begin flow of main program: define variables: accept inputs

time1 = linspace(1, 1024, 1024);
time2 = linspace(1, 80, 80);
A_cxr = 1.;
m = input("\n  ENTER MODULATION FACTOR:");
loss_dB = 0.;

# develop carrier waveform

v_cxr = A_cxr * cos(2 * pi * time1 / 32);

# develop the audio waveform: the usual coefficient
# is multiplied by 2 here so that inspection of
# one side of the two-sided spectrum will yield the
# same amplitude that would be displayed if a single
# sided spectrum were used

v_mod = m * sin(2 * pi * time1 / 512);

# display the input audio waveform
```

```
title "TIME DOMAIN TRANSMITTER AUDIO INPUT";
xlabel "TIME: 1.95 us PER SAMPLE";
ylabel "AMPLITUDE";
grid
axis([0, 512, -1.1 * abs(max(v_mod)), 1.1 * abs(max(v_mod))]);
plot(time1, v_mod, "-");
pause;

# modulate the carrier with the audio in an AM modulator that
# will not overmodulate, producing a DSBTC signal

v_cxr_am = v_cxr .* (1 + v_mod);

# modulate the carrier with the audio in an AM modulator that
# simulates a plate or collector modulator that cannot handle
# reverse current and that can therefore be overmodulated

#for n = 1:1024
#   if((1 + v_mod(n)) >= 0.)
#     v_cxr_am(n) = v_cxr(n) .* (1 + v_mod(n));
#   else
#     v_cxr_am(n) = 0.;
#   endif
#endfor

# display the time domain AM signal

title "TIME DOMAIN AM SIGNAL AT TRANSMITTER OUTPUT";
xlabel "TIME: 1.95 us PER SAMPLE";
ylabel "AMPLITUDE";
grid
axis([0, 512, -1.1 * abs(max(v_cxr_am)), 1.1 * abs(max(v_cxr_am))]);
plot(time1, v_cxr_am, "-");
pause;

# transform the signal into the frequency domain

s_cxr_am = fft(v_cxr_am) / 512;

# display the frequency domain AM signal

title "FREQUENCY SPECTRUM OF AM SIGNAL AT TRANSMITTER OUTPUT";
xlabel "FREQUENCY IN kHz";
ylabel "AMPLITUDE";
grid
```

```
axis([0, 30, 0, 1.1 * abs(max(trunq(s_cxr_am)))]);
plot(w1, abs(trunq(s_cxr_am)), "^");
pause;

# detect the signal by half-wave rectifying it as in a crystal
# set: no selectivity is provided here: it is assumed that the
# desired signal is the only one present

#for n = 1:1024
#   if (v_cxr_am(n) < 0)
#     v_cxr_rect(n) = 0;
#   else
#     v_cxr_rect(n) = v_cxr_am(n);
#   endif
#endfor

# detect the signal by full-wave rectifying it:

for n = 1:1024
  v_cxr_rect(n) = abs(v_cxr_am(n));
endfor

title "TIME DOMAIN RECTIFIED AM SIGNAL IN RECEIVER";
xlabel "TIME: 1.95 us PER SAMPLE";
ylabel "AMPLITUDE";
grid
axis([0, 512, -1.1 * abs(max(v_cxr_rect)), 1.1 * abs(max(v_cxr_rect))]);
plot(time1, v_cxr_rect, "-");
pause;

# transform the signal back into the frequency domain

s_cxr_rect = fft(v_cxr_rect) / 512.;

# display the rectified signal

title "FREQUENCY SPECTRUM OF RECTIFIED AM SIGNAL IN RECEIVER";
xlabel "FREQUENCY IN kHz";
ylabel "AMPLITUDE";
grid
axis([0, 40, 0, 1.1 * abs(max(trunq(s_cxr_rect)))]);
plot(w1, abs(trunq(s_cxr_rect)), "^");
pause;
```

```
# low pass filter the rectified signal to yield the audio output:
# this is often done with a capacitor across the headphones in a
# crystal set: note that the low-pass spectra in both halves of
# the double-sided spectrum remain after filtering

for n = 9:1015
  s_cxr_rect(n) = 0;
endfor

# pass the signal through a series "capacitor" to eliminate the DC
# component: this is often not done in a simple receiver, but it
# aids display of the received signal here

s_cxr_rect(1) = 0;

# transform the audio signal back into the time domain

v_audio = 512. * ifft(s_cxr_rect);

# display the received audio signal

title "TIME DOMAIN RECEIVER AUDIO OUTPUT";
xlabel "TIME: 1.95 us PER SAMPLE";
ylabel "AMPLITUDE";
grid
axis([0, 512, -1.1 * abs(max(v_audio)), 1.1 * abs(max(v_audio))]);
plot(time1, v_audio, "-");
pause;

# end program

exit;
```
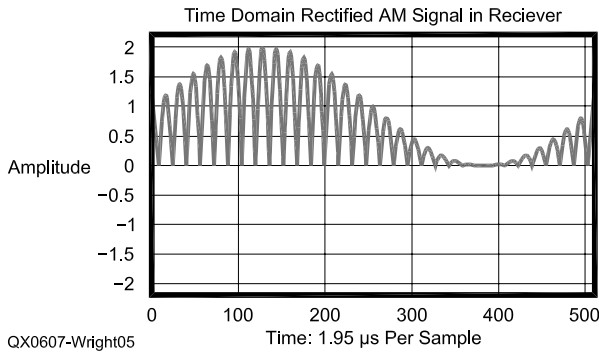
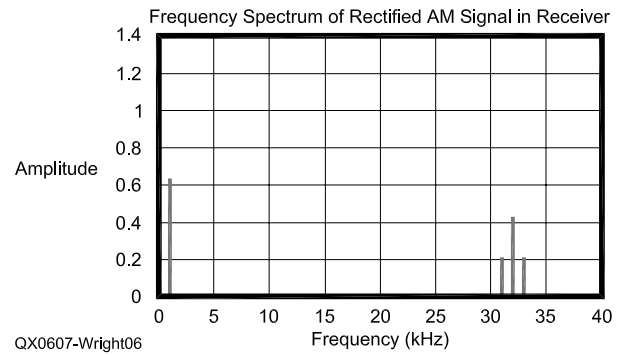Figure 5 — The rectified signal in a receiver, in the time domain.

Figure 6— The rectified signal in a receiver, in the frequency domain.
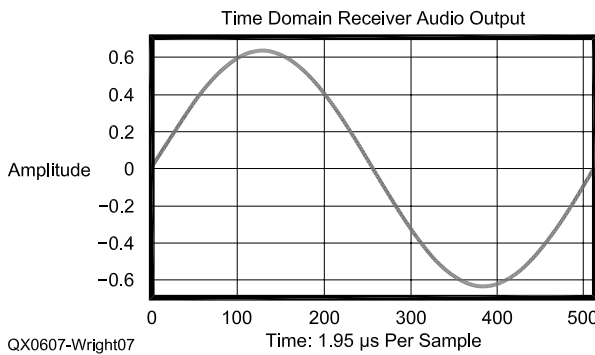

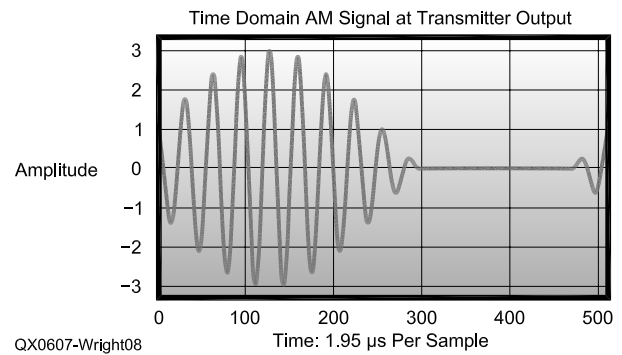Figure 7 — Recovered 100% audio signal, shown in the time domain.

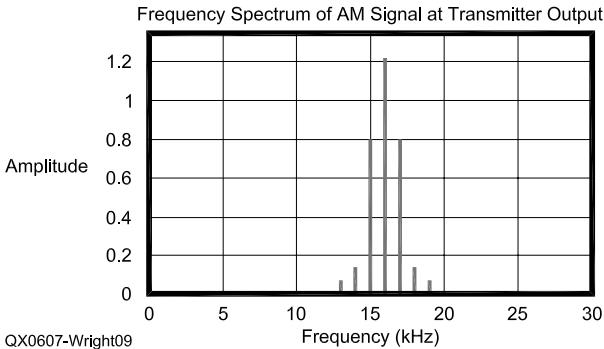Figure 8 — A 200% modulated signal, in the time domain.

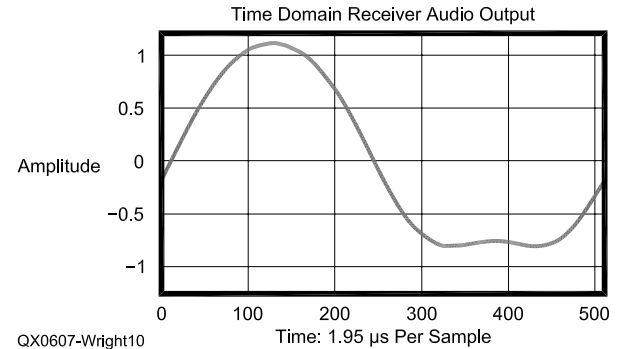Figure 9 — A 200% modulated signal, in the frequency domain.

Figure 10 — Recovered 200% audio signal, shown in the time domain.

The receiver will be a simple rectifying detector as described in Figure 11.9 of *The ARRL Handbook*, very much like a crystal set, but also representing many AM detectors in receivers that do not include product detectors.[4]

The *Octave* code for our model is listed in Table 1. The first line in the file, "#! /usr/bin/octave -qf," should be used only if you are running *Octave* under *Linux*, and you should adjust the path in that line to reflect your path to the *Octave* executable. Note that there are two modulation schemes encoded. One simulates a plate modulator or other modulator in which reverse dc cannot flow. The other, commented out in Table 1, models a modulation scheme such as that used in a typical SSB transmitter.

A balanced modulator is used to generate the modulated signal without a carrier, and the modulator is unbalanced to allow some carrier to be transmitted. In such a system, overmodulation such as that encountered in a plate or collector modulator doesn't exist.

For a modulation index of 1, the audio input — a sine wave at 1 kHz — is shown in Figure 2. The carrier is a similar sine wave at 16 kHz. The amplitude modulated signal at the output of the transmitter for a modulation index of 1 (100% modulation: page 11.5 of the *ARRL Handbook*) is shown in Figure 3. The carrier is at 16 kHz and there are two sidebands at 15 kHz and 17 kHz. The frequency spectrum of the modulated signal is shown in

Figure 4. Figure 5 shows the signal in the receiver after rectification. The spectrum of the rectified signal is shown in Figure 6. The recovered audio signal, after low-pass filtering, is shown in Figure 7. Ideally, it will be identical in shape to the input audio signal at the transmitter although it may differ in amplitude at any particular point in the circuit.

Notice from Table 1 that the code, as presented, uses a full-wave rectifier as a detector. Commented out is code for a half-wave rectifier. You can implement that by uncommenting it and commenting out the full-wave code. If you do that, you will see that the spectral components at the output of the full-wave rectifier are at double the fre-
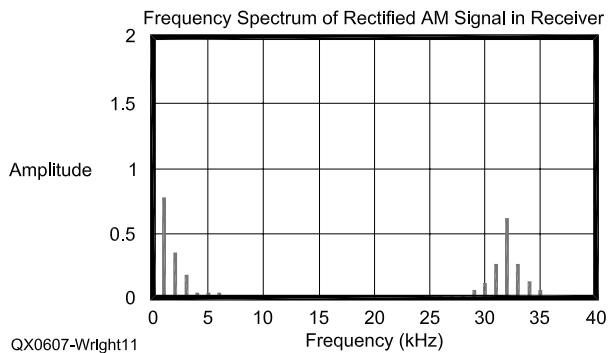
Figure 11 — The spectrum of a rectified signal in receiver, with a non-plate-modulated transmitter at 200% modulation.
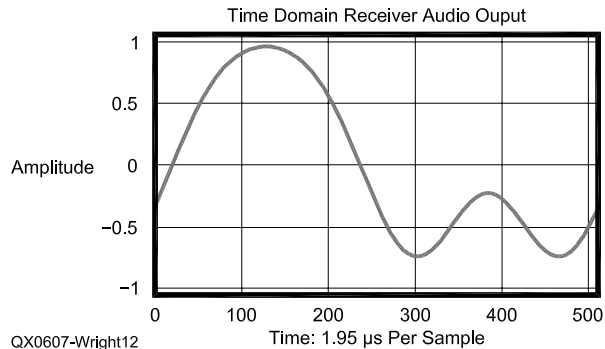

Figure 12 — Audio output of a receiver with non-plate-modulated transmitter at 200% modulation.

quencies of the corresponding components produced by the half-wave rectifier.

Now let's overmodulate the signal to see what happens. The audio input signal remains the same, but we will add some audio gain prior to the modulator in our hypothetical system to double the input voltage. We do that by specifying a modulation index of 2 when we run the *Octave* code in Table 1. That's a pretty hefty dose of overmodulation, but we'll use it to accentuate the spurious signals in our plots of the signals along the path through our model.

Figure 8, like Figure 3, shows the time domain output of the modulator. Notice, though, the long stretch of time during which the signal remains at zero amplitude. This is due to an attempt to reverse the plate current of the final amplifier by the audio waveform during a portion of its cycle. Figure 9 shows the frequency spectrum of the output signal. Note the distortion products that accompany the two normal sidebands. Figure 10 shows the time domain waveform at the output of the detector in the receiver. You can see the distortion of the negative half cycle of the sine wave.

Let's try this all again for a modulator that doesn't overmodulate on negative peaks. We can do that by commenting out the plate modulator in the code of Table 1 and uncommenting the other modulator. A snippet of code illustrating that is shown in Table 2.

If we run the new code at a modulation index of 1 or less, the results will be exactly the same as we saw when we used the other modulator. Let's try a modulation index of 2 again, though. When we do that, we see that everything looks normal in the transmitter. Unlike the spectrum with distortion products in Figure 9, we see a nice clean signal with a carrier and two sidebands, all of the same amplitude in this case.

Now let's proceed on through the code to see what happens in the receiver. Although we might infer some trouble from the appearance of the envelope of the modulated signal in the transmitter or the envelope of the rectified signal in the receiver, the first direct indication that something is awry is found

Table 2
Code Snippet Adjusted to Show Other Modulator in Effect

```
< code prior to this point omitted >


# modulate the carrier with the audio in an AM modulator that

# will not overmodulate, producing a DSBTC signal


v_cxr_am = v_cxr .* (1 + v_mod);


# modulate the carrier with the audio in an AM modulator that

# simulates a plate or collector modulator that cannot handle

# reverse current and that can therefore be overmodulated


#for n = 1:1024
#   if((1 + v_mod(n)) >= 0.)
#     v_cxr_am(n) = v_cxr(n) .* (1 + v_mod(n));
#   else
#     v_cxr_am(n) = 0.;
#   endif
#endfor


< code after this point omitted >
```

when we examine the frequency domain of the rectified signal in Figure 11 and see some interesting distortion products along with the desired signal. An examination of the recovered audio signal at the output of the receiver, shown in Figure 12, reveals even more distortion of the waveform than we saw with the plate modulated transmitter.

What happened here? Our transmitter is clean, but now our receiver is having trouble with overmodulation. In order to demodulate an AM signal without distortion using envelope detection, the peak carrier amplitude must be equal to or greater than the peak amplitude of the modulating envelope that is encoded in the sidebands.[5]

The distortion we see here may occur in listening to amateur, domestic broadcast, or shortwave broadcast AM signals when selective fading occurs and drops the carrier to a level lower than would correspond to 100%

modulation given the levels of the received sidebands.

Another application of what we have learned here involves SSB transceivers and the increasing popularity of AM on the amateur bands. A form of AM called "single-sideband, full carrier AM," "AM equivalent" (AME) or "compatible AM" (CAM) uses a single sideband signal with transmitted carrier. (See *The ARRL Handbook*, p 11.7, referenced in Note 4.) This represents a practical case for amateurs because SSB transceivers that are operated on AM by unbalancing the balanced modulator to inject carrier often operate in this manner.[6]

You can generate a single-sideband, full carrier AM signal in the *Octave* AM system model by filtering out one sideband after modulation. Adding the code shown in Table 3 will eliminate the upper sideband. When you do that, try the code again. Note

that we have filtered both the positive and negative frequencies, something that you need to keep in mind as you add filters or other frequency-selective devices to a model.

You may be surprised to find that, in the full carrier AM model, even when the modulation index is lower than one, distortion products remain. While a double-sideband, transmitted carrier (DSBTC or AM) signal can be envelope detected without distortion when the peak amplitude of the carrier is equal to or greater than the peak amplitude of the encoded modulating signal, such is not the case for a single-sideband, full carrier AM signal. You can reduce the distortion to a very low level, though, by making the carrier amplitude much higher than the amplitude of the encoded modulation.[5]

This may have practical implications if you have an SSB transceiver and would like to operate it in the AM mode. One of the attractions of AM is the excellent audio quality obtainable. We can transmit a distortion-free DSBTC signal from an SSB transmitter.[6] An older receiver, though, which might be used in a legacy AM station, may feature only envelope detection, and may therefore not produce an audio output from a signal generated by an SSB transceiver with the same quality that would be perceived with a signal from an AM transmitter.

I haven't been on AM for over forty years (I'm tempted, though), so I can't attest to whether or not this is a real problem. Look, though, at what we have learned and stirred up just by using *Octave* to poke around in a really simple system model.

You can enhance this model to simulate SSB transmission with a product detector at the receiver. You can add noise (use the *Octave* function *randn( )*), attenuation, and distortion to the channel between the transmitter and receiver. You can simulate a digital signal by restricting the amplitudes to discrete samples and you can experiment with some of the principles described in Chapter 16, *DSP and Software Radio Design,* of *The ARRL Handbook.* There are a lot of things you can do with *Octave* or with other math programs and we've tried to demonstrate some of them here.

In this article and in *Octave for Signal Analysis,* frequencies were carefully chosen to prevent various impairments that may be encountered in representing continuous signals by a sequence of discrete samples. Aliasing and leakage are described in Chapter 16 of *The ARRL Handbook.* The picket fence effect is described on page 544 of Lathi.[7] In simulating more complex circuits or in simulating real software or hardware, you may not be able to easily avoid such impairments. Aliasing can be avoided by making sure that the sampling rate associated with your model is more than twice the maximum frequency of any components in

Table 3
Code Snippet to Produce Single-Sideband, Full Carrier AM

```
< code prior to this point omitted >


# transform the signal into the frequency domain


s_cxr_am = fft(v_cxr_am) / 512;


# filter the upper sideband, producing a lower-sideband,
# transmitted carrier signal


for n = 34:39
    s_cxr_am(n) = 0.;
endfor
for n = 987:992
    s_cxr_am(n) = 0.;
endfor
v_cxr_am = ifft(s_cxr_am);


# display the frequency domain AM signal


title "FREQUENCY SPECTRUM OF AM SIGNAL AT TRANSMITTER OUTPUT";
xlabel "FREQUENCY IN kHz";
ylabel "AMPLITUDE";
grid
axis([0, 30, 0, 1.1 * abs(max(trunq(s_cxr_am)))]);
plot(w1, abs(trunq(s_cxr_am)), "^");
pause;


< code after this point omitted >
```

your signal. Leakage can be minimized by using the windowing functions shown on page 16.29 of *The ARRL Handbook.* Picket fence effect can be minimized by using enough samples so the frequency resolution is sufficient for your purposes.

Notes
[1]M. Wright, W6PAP, "*Octave* — Calculations for Amateurs*,"* May/June 2005 *QEX,* pp 48-50.
[2]This assumes that you are willing to comply with the GNU Public License, which you can read at **www.gnu.org** or **www.octave. org**. See **www.octave.org** to obtain a copy of *Octave* for your system.
[3]M. Wright, W6PAP, "*Octave* for Signal Analysis*,"* July/August 2005 *QEX,* pp 57-61.
[4]*The ARRL Handbook for Radio Communications — 2006,* The American Radio Relay League, 2005. *The ARRL Handbook* is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 9485. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.
[5]B. P. Lathi, *Modern Digital and Analog Communication Systems, Third Edition,* Oxford University Press, 1998, p 178.
[6]*Operation and Maintenance — Swan Cygnet Model 270, Revision 1,* Swan Electronics, 1970.
[7]B. P. Lathi, *Linear Systems and Signals,* Berkeley-Cambridge Press, 1992.

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. ▢EX▬

An Innovative 2-kW Linear Tube Amplifier

Can a tube full-legal-limit amplifier be small and light?

Saulo Quaggio, PY2KO

This prototype was built for testing a new concept in linear amplifiers, which could be named AB/F class. It is not actually a new operating class, but has characteristics of the well known class AB and class F applied to the same amplifier. It is composed of a bunch of old and new technologies, resulting in a very small and light amplifier able to operate in AM, CW and SSB at legal limit power.

The idea was to get linear performance from a non-linear high efficiency amplifier. To achieve it I adopted an envelope elimination and restoration approach: The incoming signal is demodulated and the original envelope amplified by a high efficiency, high speed switching supply that modulates the feeding of output stage power.

The incoming RF signal is also applied to the tubes' control grids. At low power levels the grid polarization circuit keeps the amplifier at class AB. As the power gets higher, the tubes are driven into saturation and the output power is controlled by the rising voltage of the modulating power supply.

Why tubes and not semiconductors? Here are some reasons:

- Relative fragility of power FETs in high power RF circuits. Tubes are much more resistant to mistuning and unpredictable conditions that may occur during development.
- Possibility of applying this project to more powerful commercial amplifiers.
- Smaller variable capacitors at tuning output circuits, due to higher impedance levels.
- Tubes are fun! It's amazing to see four small TV tubes effortlessly pumping out RF over the legal limit. Old PL509 TV sweep tubes were built to operate as switches, so they perform very well in this saturated class, achieving high anode efficiency.

Au Prof alceu Maynard Araujo 153 uj34
CEP 047-26160
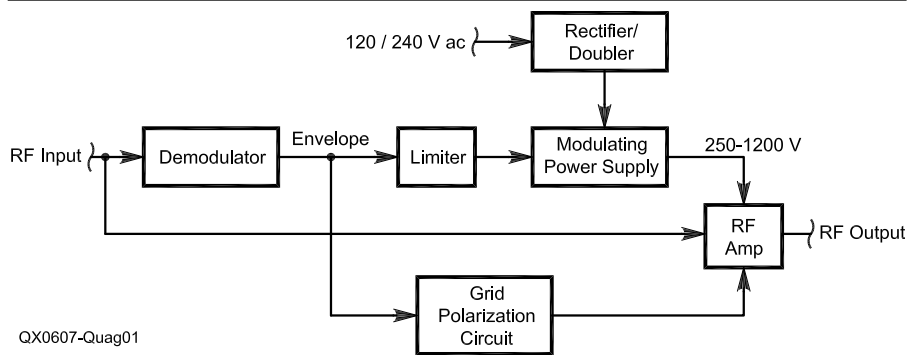Sao Paolo, Brazil
saulo@auttran.com.br

Figure 1 — Amplifier block diagram.

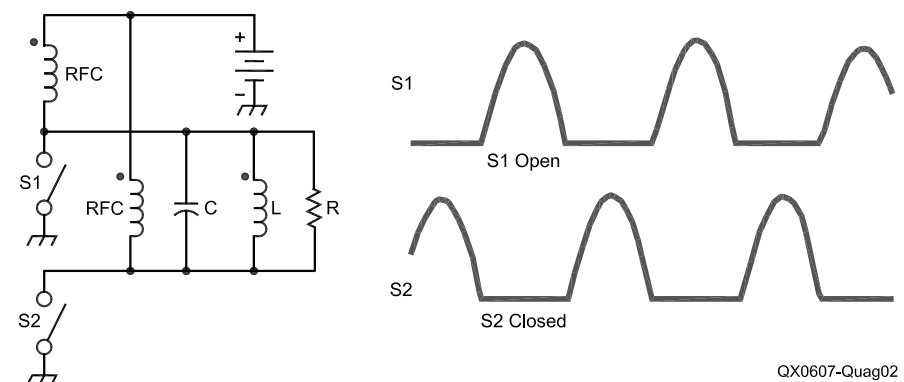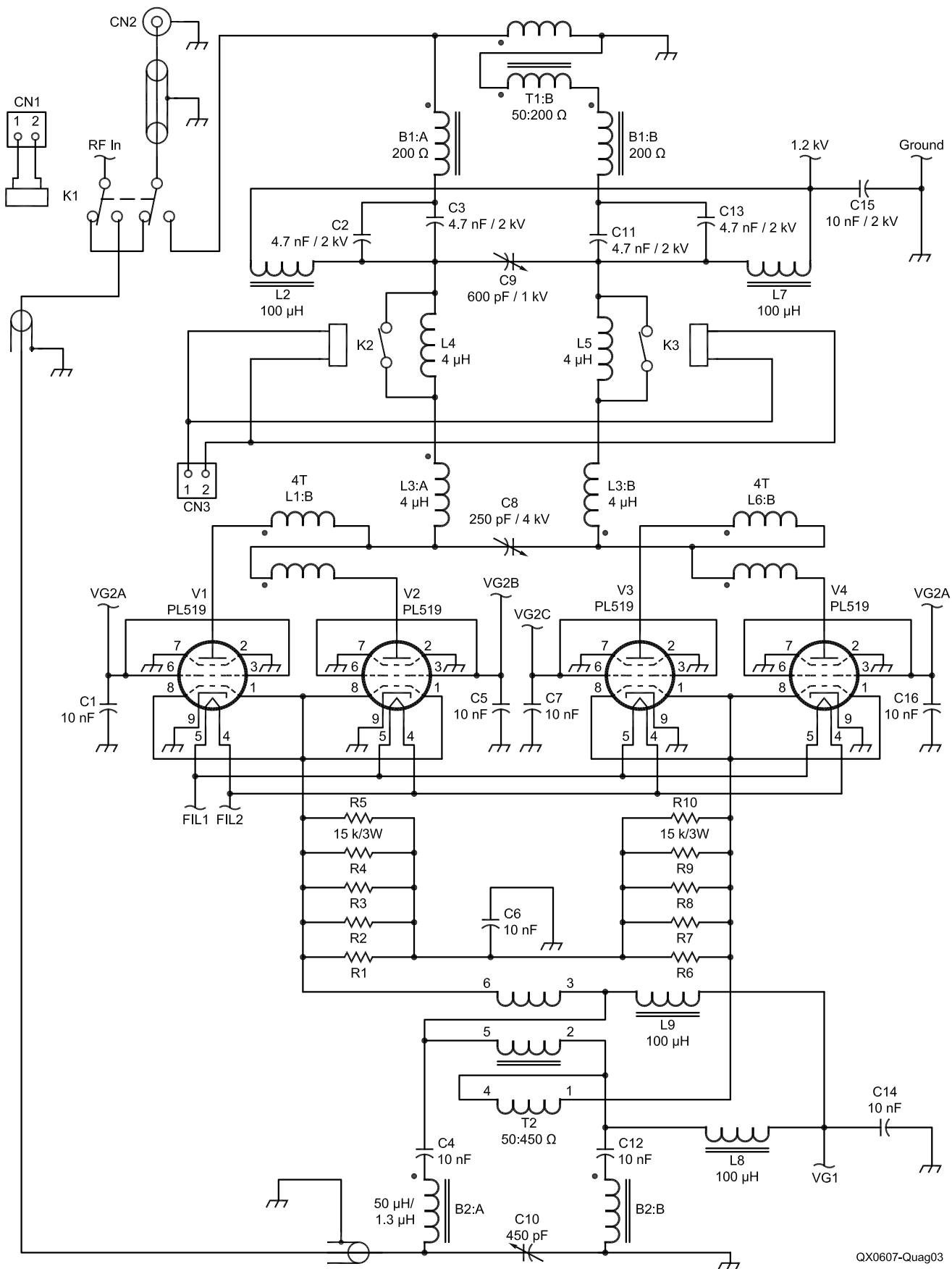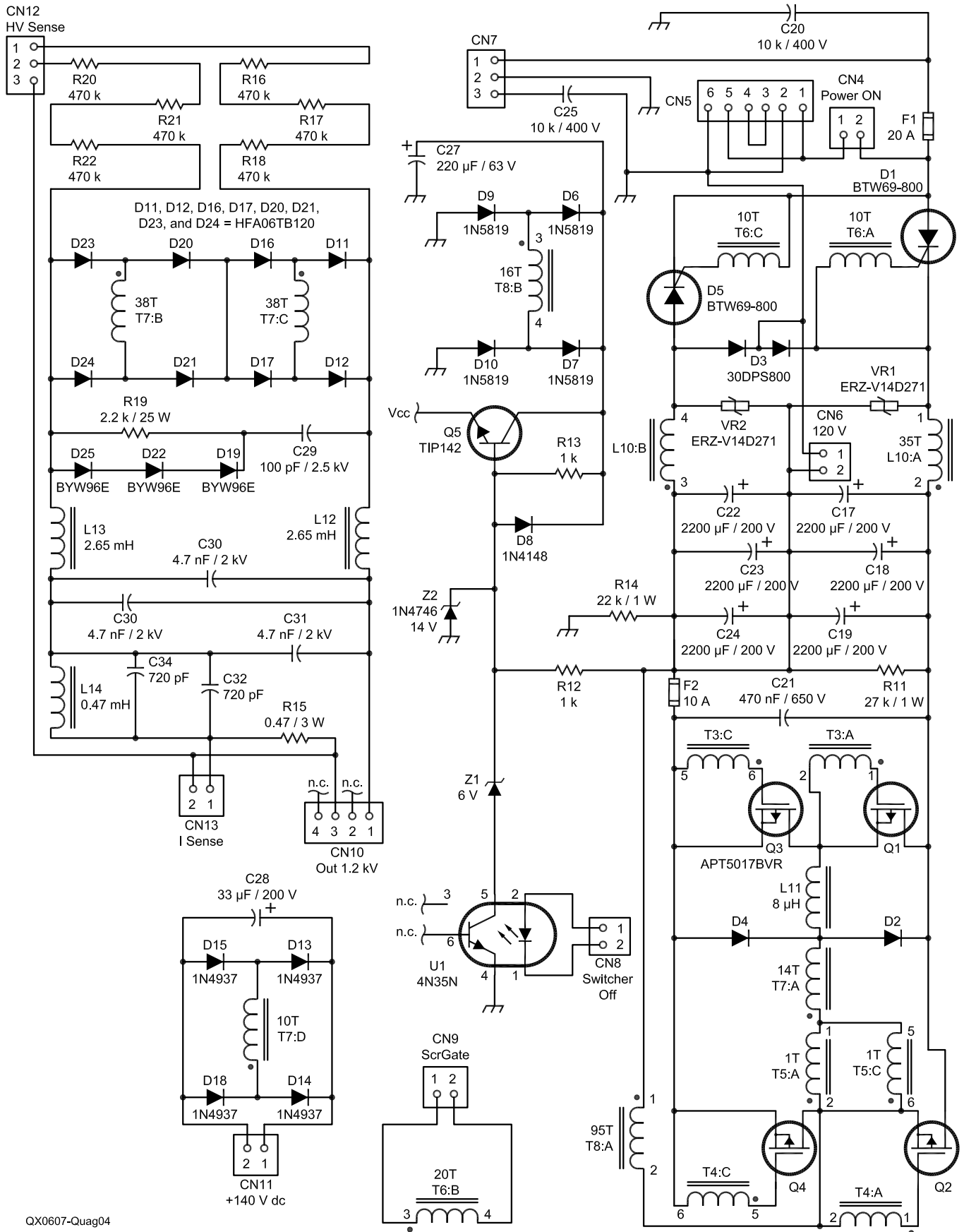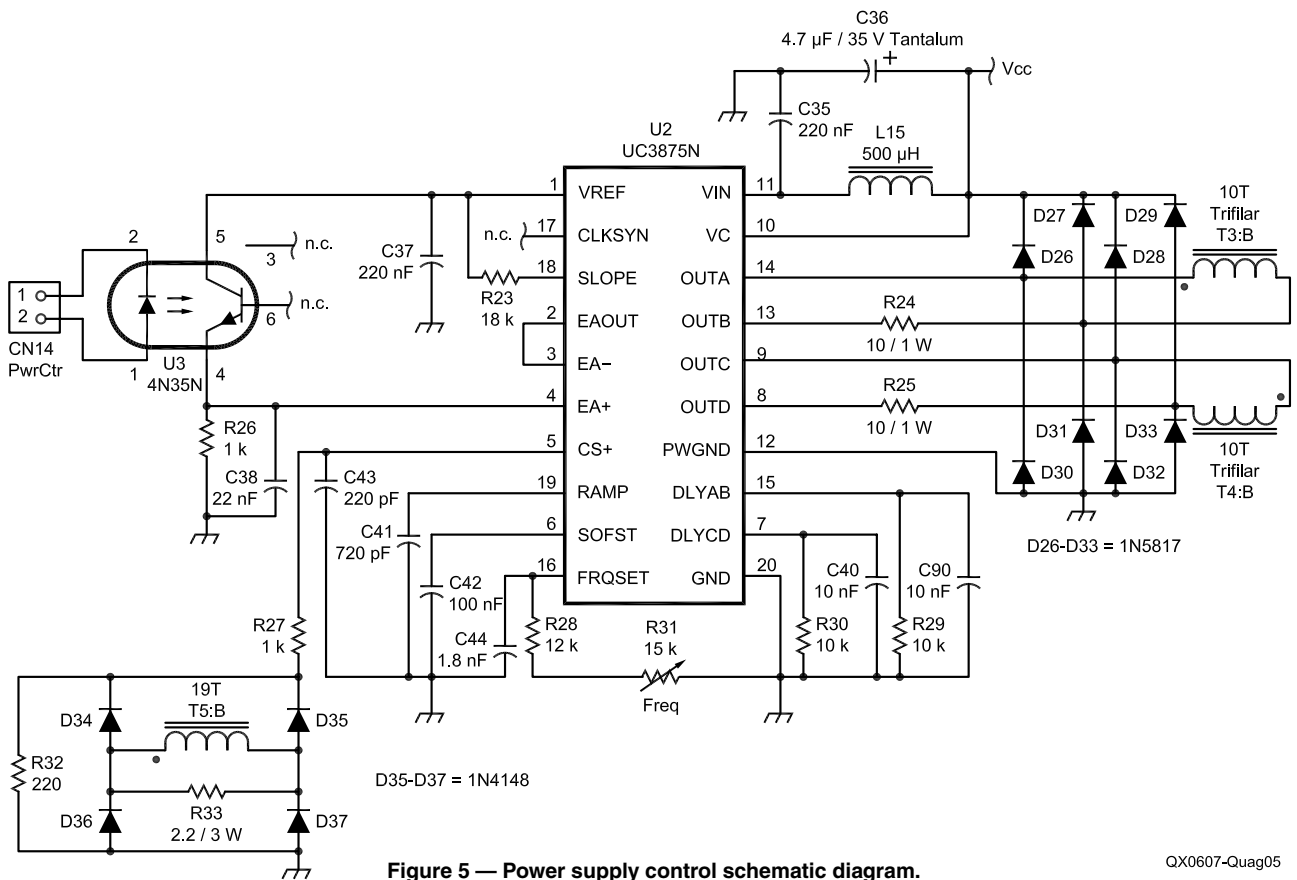
Figure 2 — Basic waveforms.

Figure 3 — RF amplifier schematic diagram.

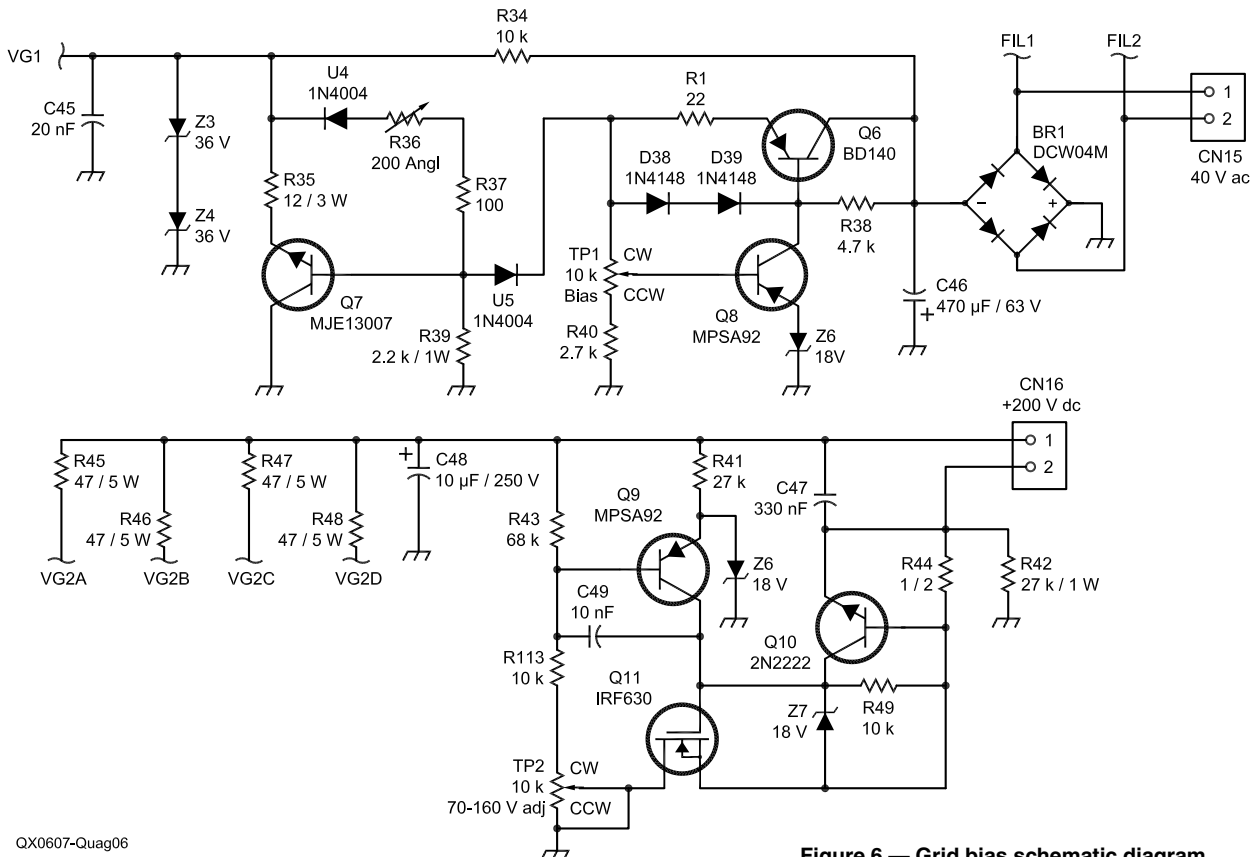**Figure 4 — Power supply schematic diagram.**

Figure 5 — Power supply control schematic diagram.

QX0607-Quag05

Figure 6 — Grid bias schematic diagram.

QX0607-Quag06
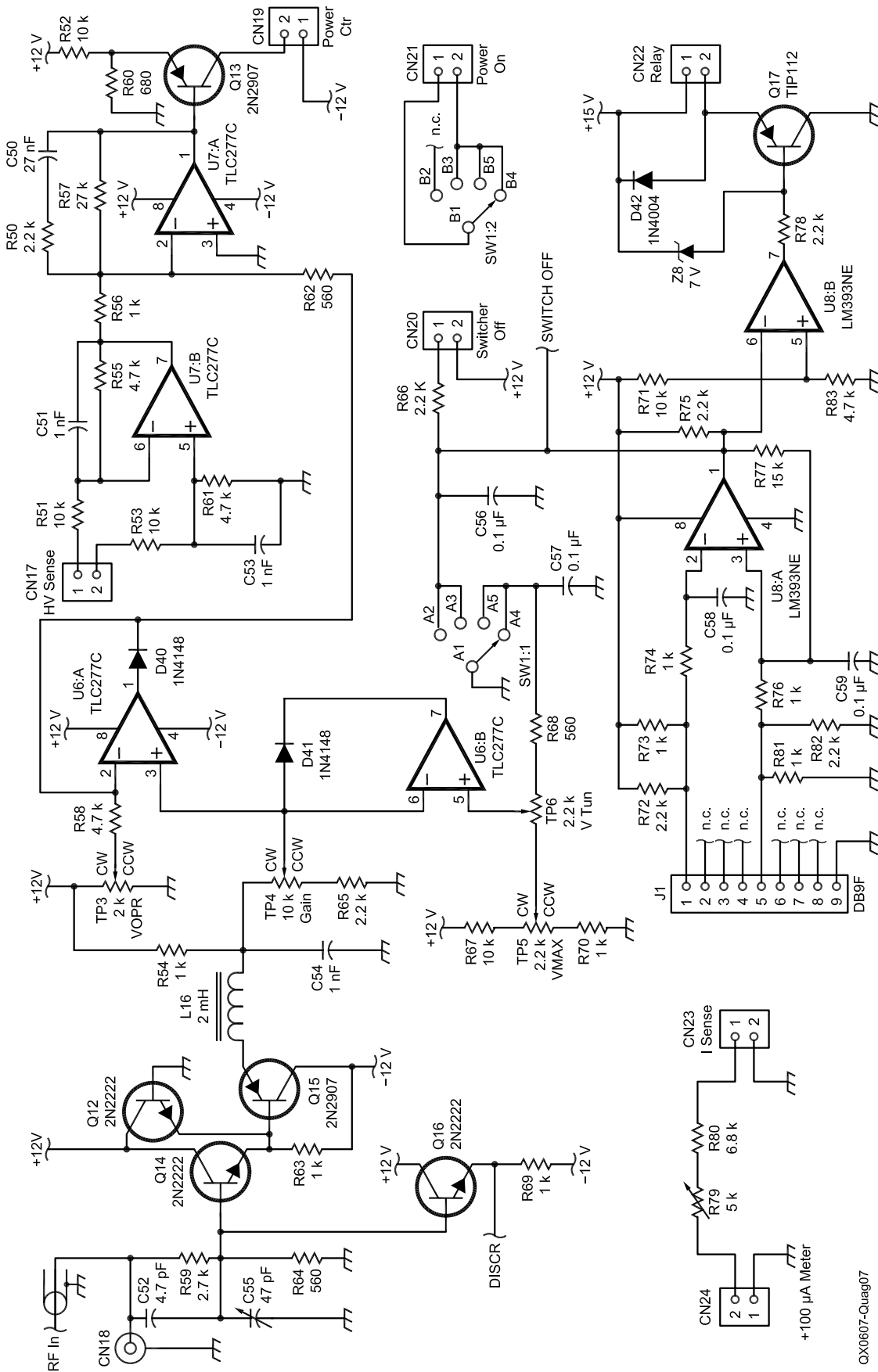
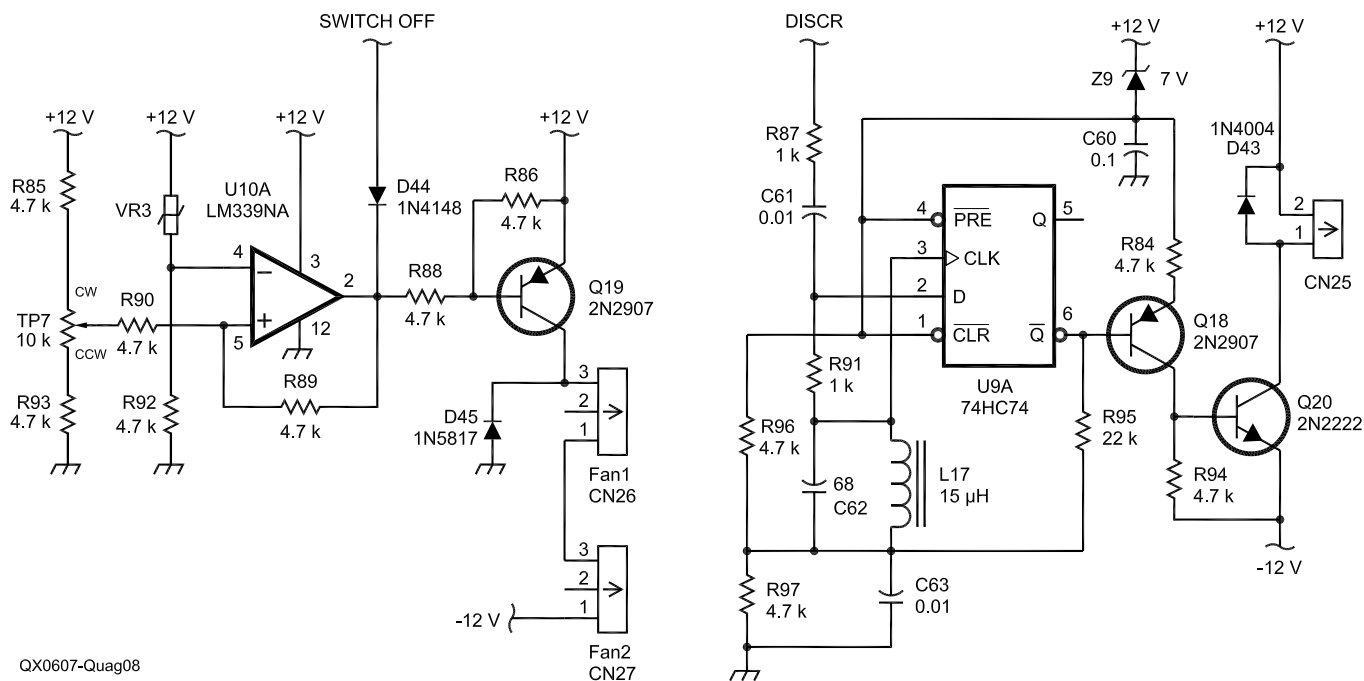**Figure 7 — Control schematic diagram.**

QX0607-Quag07

Figure 8 — Thermal protection schematic diagram.

QX0607-Quag08

(Continued from page 19.)

Refer to the block diagram of Figure 1. The amplifier is a harmonically controlled push-pull arrangement operating in a mode known as *inverse class F*. In the frequency domain, the symmetric resonant load presents an open circuit to odd harmonics, a short circuit to even harmonics and is resistive to the fundamental frequency.

Another way to explain how it works in the time domain is to consider the tubes as switches grounding alternatively the two sides of a resonant load consisting of a resistor, a capacitor and an inductor in parallel. Across each switch is a half sinusoid voltage pulse when opened, and zero volts when closed. The two halves combine at the load in a full sine wave. Two RF chokes feed dc power to the circuit. Note that the load must be isolated from ground.

Basic Waveforms

Refer to the basic waveforms shown in Figure 2. Theoretically, this mode allows 100% efficiency, since there is no voltage and current simultaneously at the switches; but because of parasitic capacitances, resistive losses and minimum anode voltage, the actual plate efficiency runs around 85%.

To ensure correct coupling, a differential adjustable pi circuit lowers the impedance from 1800 Ω at the tube side to 200 Ω, and at the same time provides the resonance needed. A 200 Ω transmission line "unun" isolates

You can see the small size by comparing the amplifier to a Bird model 43 wattmeter.

the pi circuit from ground, and a 200 to 50 Ω balun also converts from differential to common mode.

Circuit Description

Refer to the schematic diagrams in Figures 3 through 9 during the following discussion. A side benefit of the push-pull output arrangement is that since there are virtually no even harmonics generated, a loaded Q of 5 is enough to filter out the 3rd and 5th

harmonics, permitting smaller bobbins, capacitors and commutating relays than allowed by conventional unbalanced pi circuits.

A high frequency diplexer isolates the paralleled tubes against VHF oscillations. These oscillations actually occurred, making the internal tube plate connections glow red!

The incoming RF signal is detected by a discriminator circuit centered at 5.4 MHz that actuates two high-current relays, which short the bobbins, halving the pi

inductance when operating on 40 m.

This configuration allows linear class-AB push-pull operation until the amplifier reaches roughly 100 W. From this level up the tubes start to saturate, and the current through the control grids forces the polarization level more negatively, out from class-AB. The amplifier changes smoothly from class-AB to class-F as the plate voltage rises following the input signal envelope.

The input circuit is a wideband balanced circuit, operating with no tuning from 3 to 8 MHz with reasonable SWR. The input transformer is a 1:9 transmission-line type loaded by an array of resistors paralleled to the control grids.

The Switching Power Supply

The power supply is an off-line, full-bridge, phase-controlled, quasi-resonant converter operating at 100 kHz, capable of 2500 W. Phase control means that each side of the bridge operates always with square waveforms, and the control circuit changes the relative phase between bridge sides from zero to 180°, varying the output voltage from zero to 1200 V.

The quasi-resonant approach was adopted not only to increase efficiency but also to reduce secondary ringing by slowing primary turn-on slope. Unlike conventional switchers, this one has broad response bandwidth, necessary to follow the audio envelope of the incoming reference signal. Stability compensation versus ripple attenuation limits the bandwidth to 8 kHz, enough to follow AM and SSB input signal envelopes.

The output filter is a 2nd-order 15-kHz low-pass Chebyshev, followed by a 200-kHz series trap. The commutating ripple is attenuated 60 dB by this filter.

It is interesting to note that the small capacitors at the output and the fast current limiting gives a "soft" characteristic to this power supply: A sudden short circuit at full power only fires a small spark, very unlike a conventional high voltage supply of this size!

The 240 V ac power input is rectified by an SCR-diode bridge, provided there's a soft start circuit to limit inrush current and an input choke to improve the power factor. The SCRs also act as the main power relay. The input rectifier could work as a doubler to operate 120 V ac, if you set some jumpers and have heavy duty wiring in your shack.

I did not try to shield this big switcher for RFI: When not transmitting it is held totally off. This required a small 60 Hz transformer to supply tube heaters and control circuit power, but it is worth the space occupied — there's no receiving interference at all. And while transmitting? Well, if you can keep the TV sets working during your 2 kW chattering, interference from the switcher is no problem.

Demodulator, Limiter and Control Circuits

The input active differential RF detector performs better than simple diodes, presenting very good linearity. Following the RF filter, the limiting circuit keeps the power supply above 250 V, to allow the tubes to operate in class-AB at low power. It also limits the maximum voltage, to protect the amplifier if the main ac supply goes high. Without it, the instantaneous input power can reach 3 kW, meaning danger to tubes and capacitors.

Another limiter function is to keep anode voltage below 500 V during tuning, also to protect components and to give a stable tuning reference current.
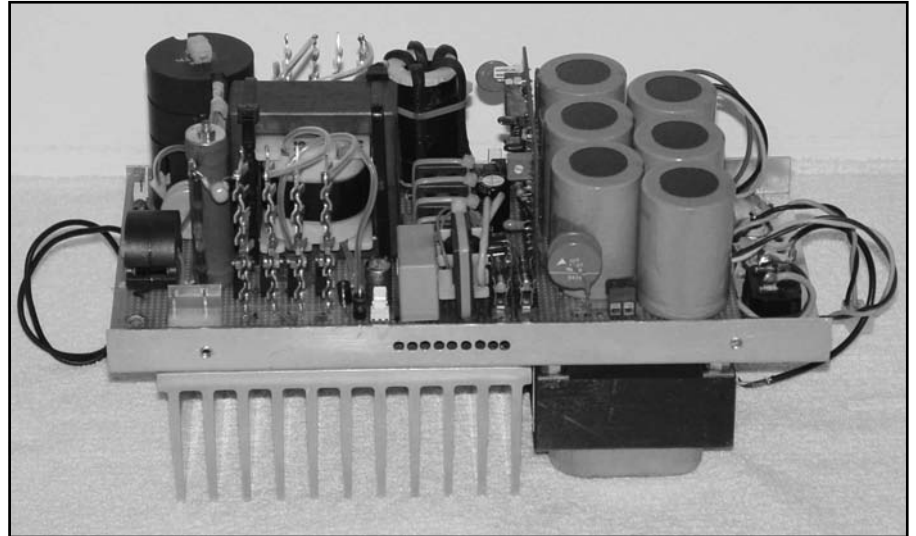
This photo shows the power supply, removed from the amplifier.
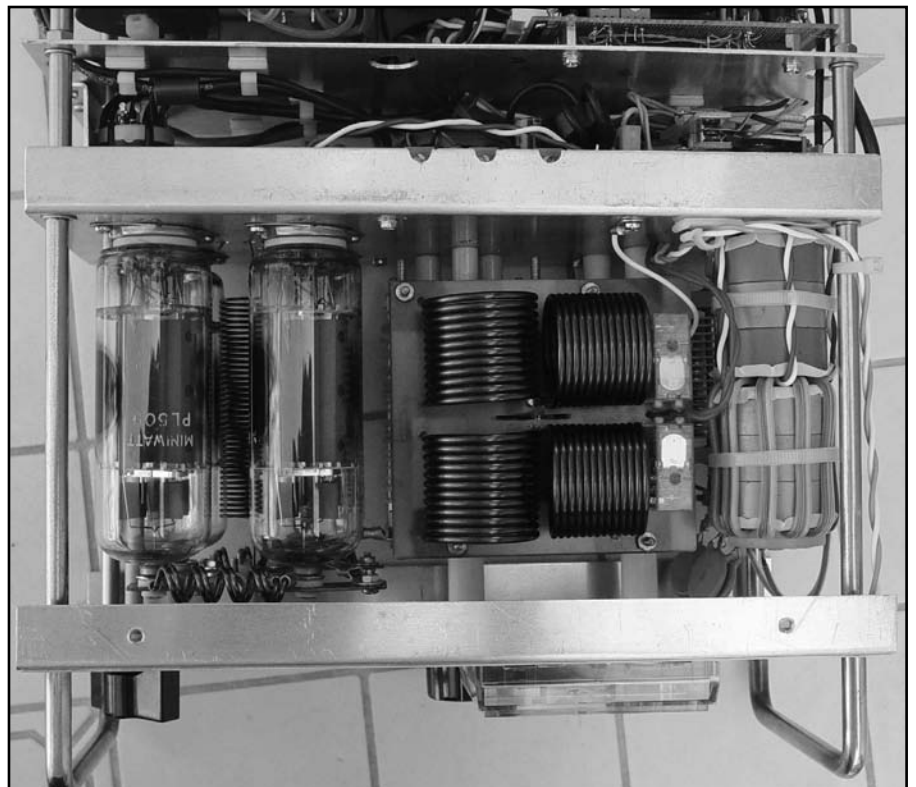
The RF deck top view: tubes, output bobbins, relays and balun stacked to output transformer. The vertical "coil" between tubes is a spring to keep them in place.
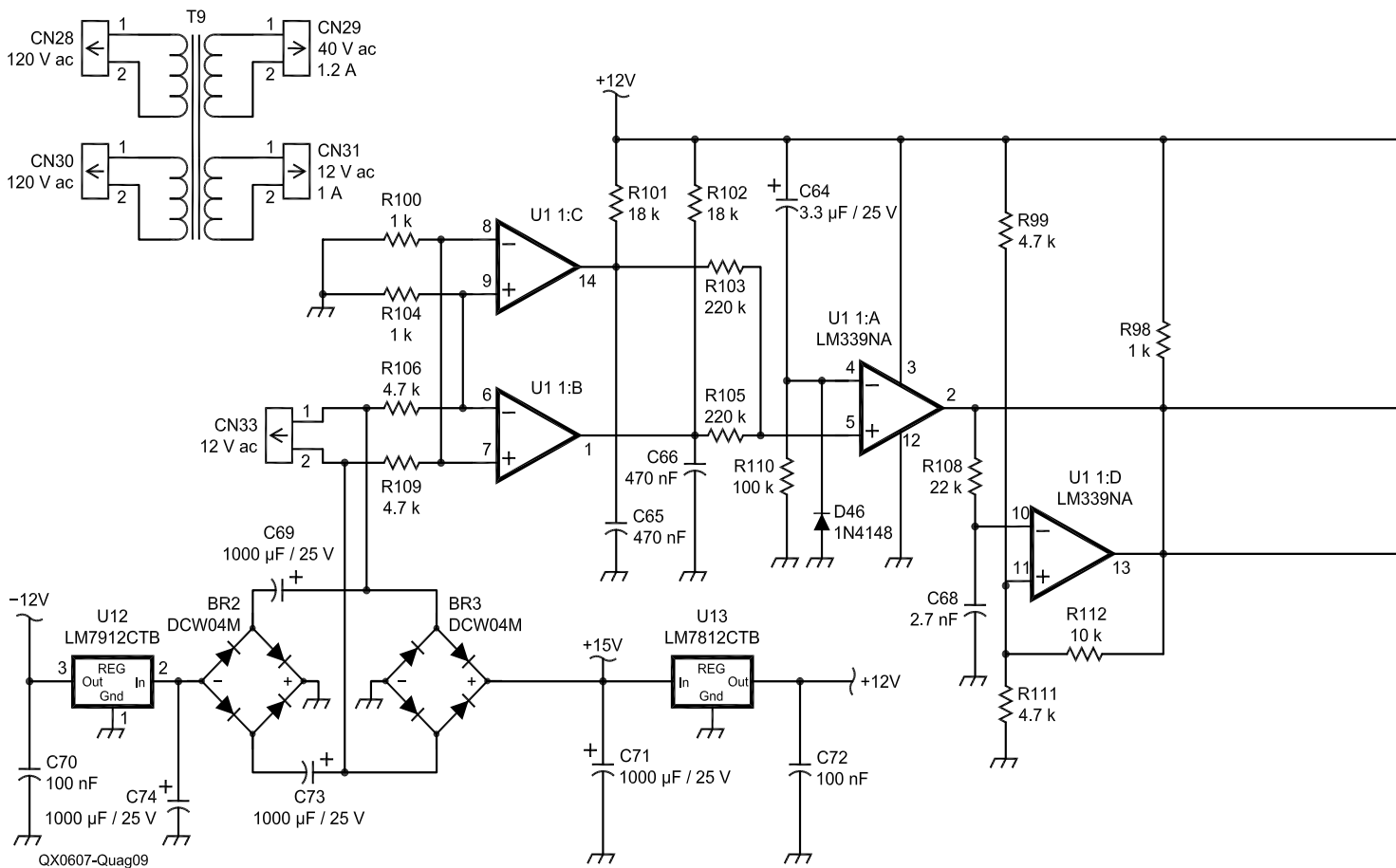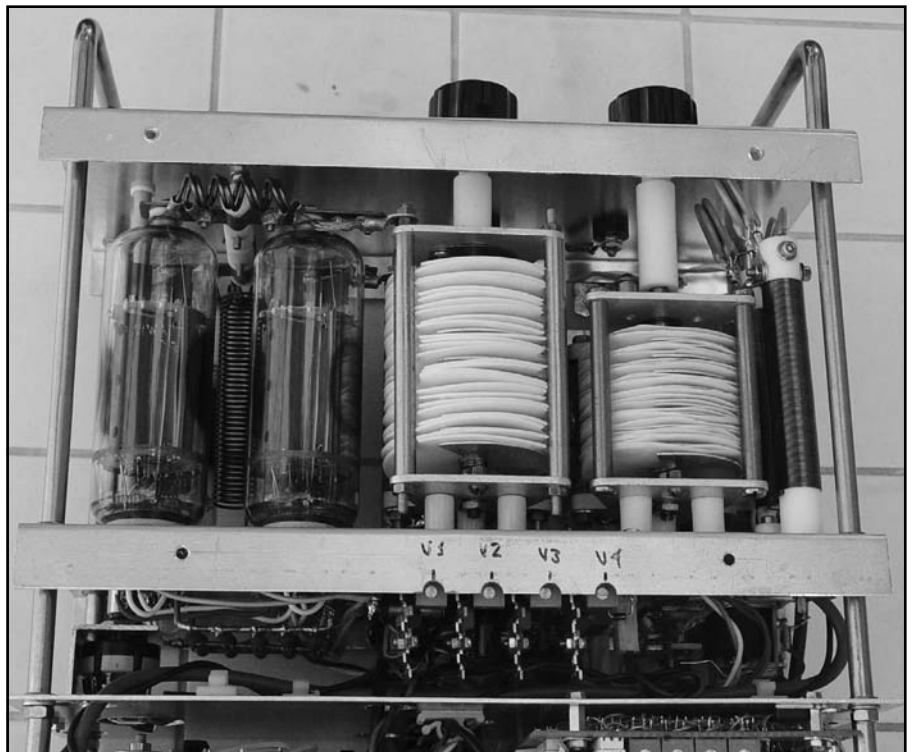
The conditioned envelope input is compared to a voltage output sample, and the error signal is amplified, compensated and applied to the power supply phase control via an optocoupler. TR switching and sequencing is performed according to the MODE switch, whose positions are:

1 — Off.

2 — Standby: Heaters and fans on, switching power supply off, RF input bypassed to output.

3 — Tuning: Anode voltage limited to 500 V, RF amplifier connected.

4 — Operating: Maximum power enabled, RF amplifier connected.

Tuning is performed by increasing drive power until saturation is reached (no further increase in plate current), at the same time setting a dip at 1.2 A adjusting alternately plate and load capacitors, as with any pi output amplifier.

The tuning procedure is the same for SSB, CW and AM. In AM the input carrier must be adjusted to read 1.2 A (approximately 500 W RF output). It must be done carefully because of the low power dissipation capabilities of the PL509. The efficiency drops quickly when it is out of tune, and if RF drive is applied care-

RF deck bottom view: tubes, variable capacitors and dc feed chokes.

Q21
2N2222

C67
10 µF / 25 V

Q22
2N2907

1

CN32
ScrGate

2

R107
10

QX0607-Quag09

Figure 9 — Auxiliary supply and soft-start schematic diagram.

lessly the thermal protection circuit can be called on to save the tubes, turning power off.

Grid Polarization Circuits

The switching supply provides unregulated 200 V dc to the screen grid series regulator. The circuit also applies negative bias to control grids, keeping the tube's quiescent current at 200 mA (class AB). When drive power reaches the point where the grids start to conduct, the bias becomes more negative due to rectification of input RF. When the tubes are in full saturation, the control grids conduct half of the RF cycle, and its voltage waveform is a negative half sinusoid, much like the plate waveform with inverse polarity.

Also like the output, these half sinusoids appear summed at the input transformer as a complete wave, so there is no input loading distortion even with broadband coupling.

Construction

The photos illustrate the construction of this project.

The whole amplifier is mounted over four aluminum decks held together by "inox" threaded bars inside inox tube separators. Four L-shaped acrylic sheets close the cabinet. This cover was made for demonstration purpose only, but as it looks good, it was kept. I am aware this is not a definitive solution, since there is a considerable RF field flowing across the operator and neighboring electronics. A cover made of identically shaped aluminum will be provided soon.
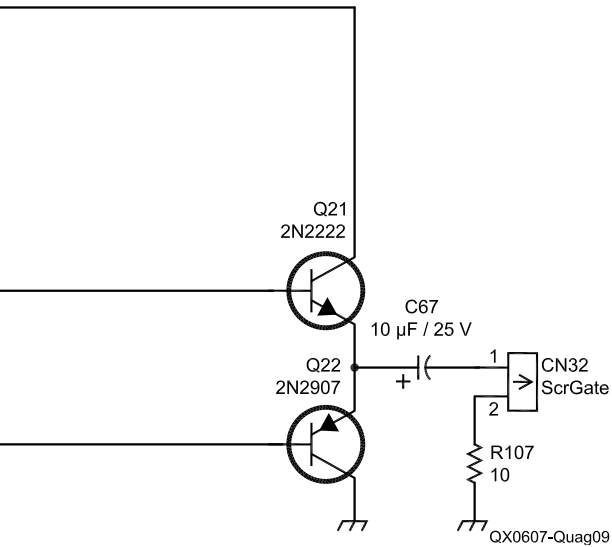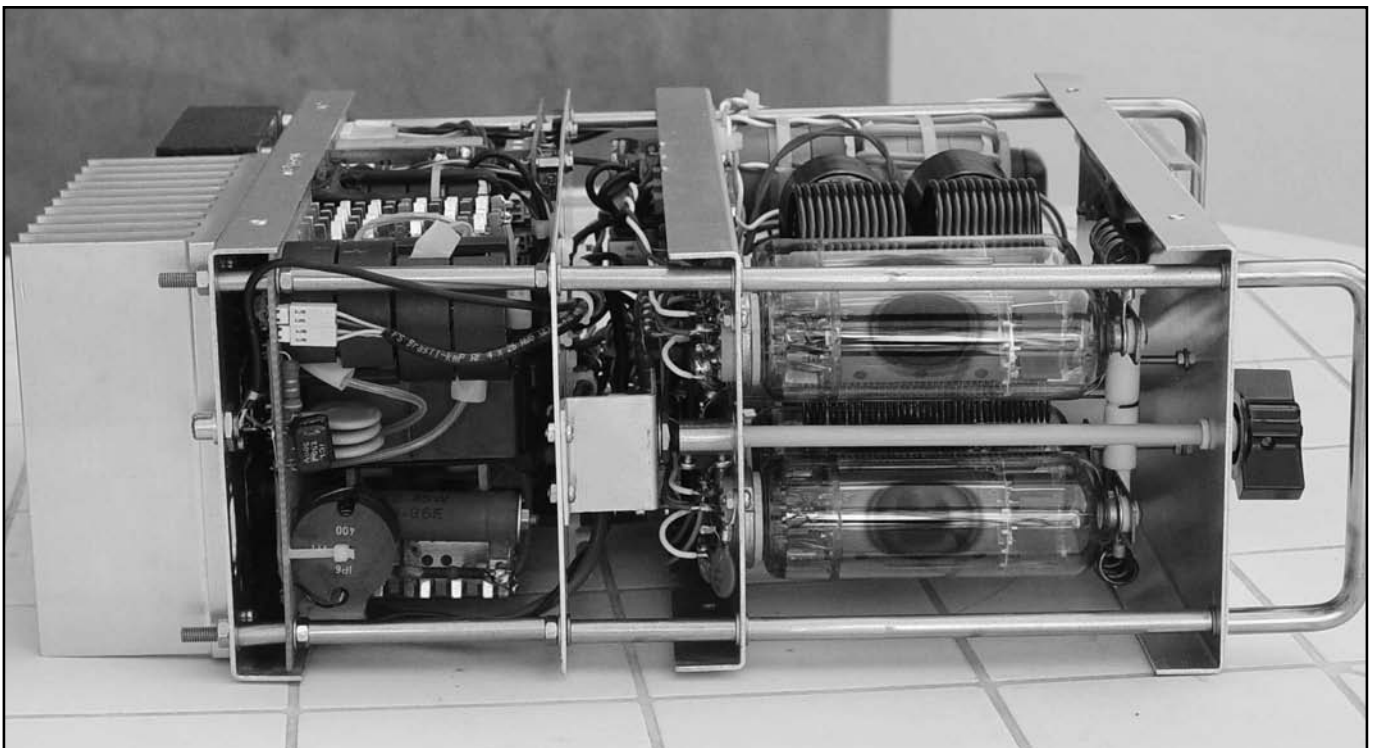

Left side view: power supply output transformer and filter, mode switch and tubes

To shrink the variable capacitors into the small space available, the aluminum plates were interleaved with Teflon discs. Because of Teflon's higher dielectric constant and electrical rigidity, it was possible to construct open capacitors similar to vacuum types regarding to size and isolation. But I am not sure if they will last long in an environment full of dust and humidity.

Two dc computer fans operating at half voltage (they are near inaudible at this power level) keep the amplifier cold even during long
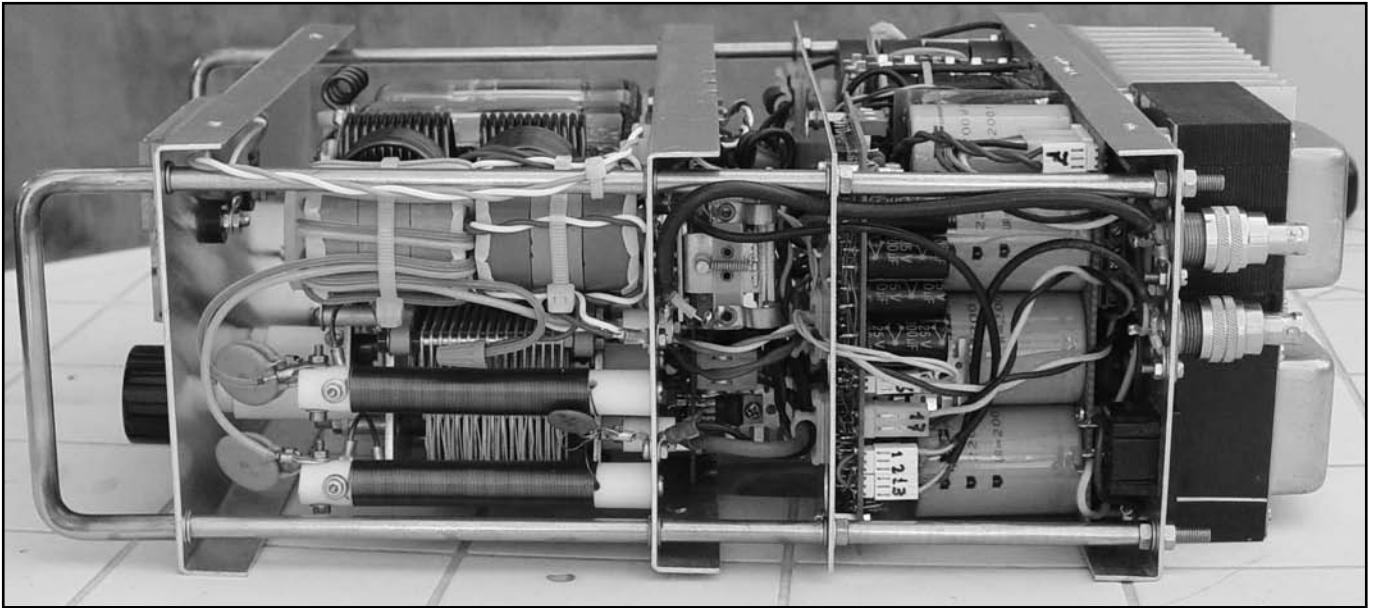
Right side view: balun and transformer, dc chokes, RF commutating relay and input storage capacitors.
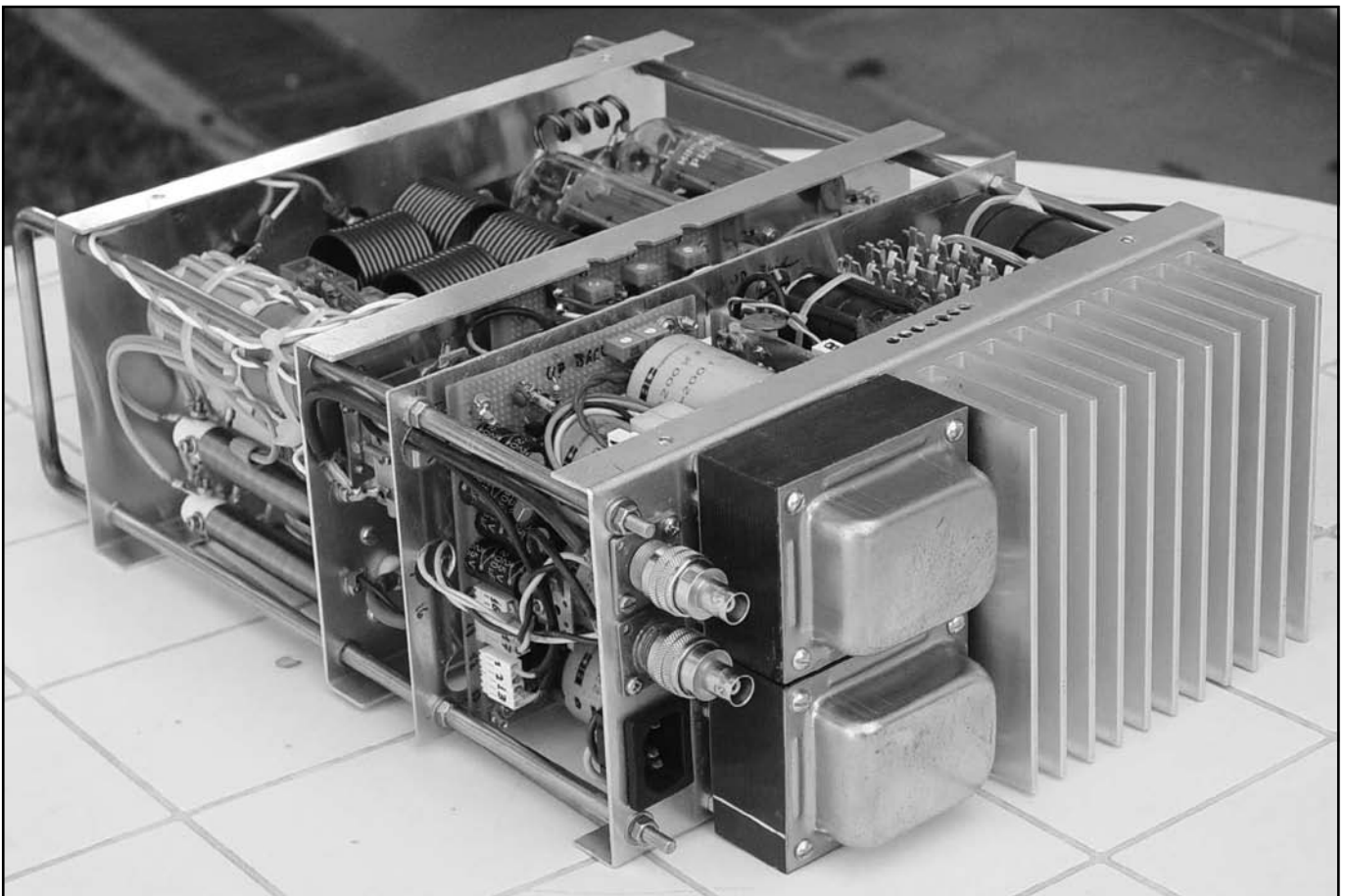
Back side: Auxiliary transformer, ac input filter choke, RF input/output connectors.

AM chats. When triggered by thermal overload, the protection circuit applies full voltage to the fans, shortening the shut-off period.

Only the critical circuit boards were made by printing (actually hand-painting and etching): The output bobbin board and power supply control/drive. Other boards were hand wired wire-wrap style but all joints were soldered.

The front panel meter is a no-barrel type to save RF deck space. All the switching supply power semiconductors are mounted over the aluminum heat sink located at the back side. The heater transformer is also mounted on the back, together with the rectifier input choke. They both are built with a silicon-iron core to reduce size and weight (see photo).

Specifications

Bandwidth: 3.5 to 3.8 MHz and 7.0 to 7.4 MHz

SWR output: 2.5:1 max.

SWR input: 1.5:1 max.

Output power
- SSB: 2000 W PEP
- CW: 1000 W, 50% duty cycle
- AM: 500 W carrier, 2000 W peak
- FM, RTTY: 500 W continuous

Power gain: 15 dB

Plate efficiency @ 1000 W: 83%

3rd-order IMD: –35 dB @ 2 kW PEP

Spurs and harmonics: – 60 dB

Dimensions: 30 × 13 × 45 cm

Weight: 8 kg

Conclusion

This prototype was a good platform to test some innovative concepts. I was surprised by its good linearity. As you can see in the oscilloscope photo, there is a visible discontinuity at crossover level, because of the limited frequency response of the switcher loop. It seems that this distortion only worsens high-order intermodulation products, making them a bit higher than measured with conventional amplifiers, but the third-order products are lower. Both are well below accepted standards in any case.

The amplifier has impressive performance, is very compact, unconditionally stable and easy to tune. There are some drawbacks on this project, however:
- It's not very easy to duplicate due to some hard to find parts as TV tubes, uncommon capacitors, old sockets, etc.
- High PL519 interelectrode capacitance precludes working at higher bands.
- The compact layout is somewhat clumsy and difficult to test and service.

But it could be a starting point to develop
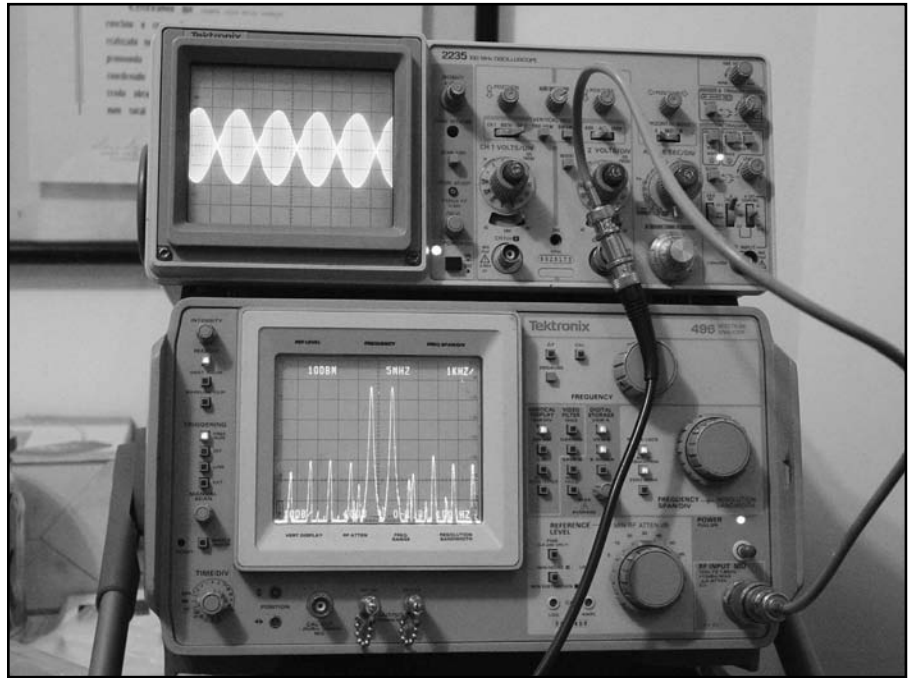
IMD at 1 kW output.

an all-HF-bands semiconductor version, microprocessor-controlled, with auto-tuning and other goodies. Another interesting project could use more powerful tubes, such as two 4CX400 tubes, which would develop a 2 kW output level continuously.

I've gotten good contacts on AM and SSB with it, and when I tell technically minded fellow hams what is inside the equipment they are hearing, they find it hard to believe: What? Plate modulated linear? Two kilowatts from TV tubes?

Saulo Quaggio holds a degree in Electrical Engineering from the University of Sao Paulo. He has been licensed as PY2KO since 1990. He has worked in several Brazilian government and commercial electronics/mechanical projects. He is dedicated to ragchewing and construction projects, and also enjoys scuba diving and flying ultralight aircraft. Saulo runs his own company (where he also works as a computer and hardware engineer) that specializes in automatic fare collection systems for public transportation.

QEX

Adventures in Curve-Fitting

By using a nonlinear transfer function to characterize a circuit, we can extend the useful range beyond the linear segment of operation.
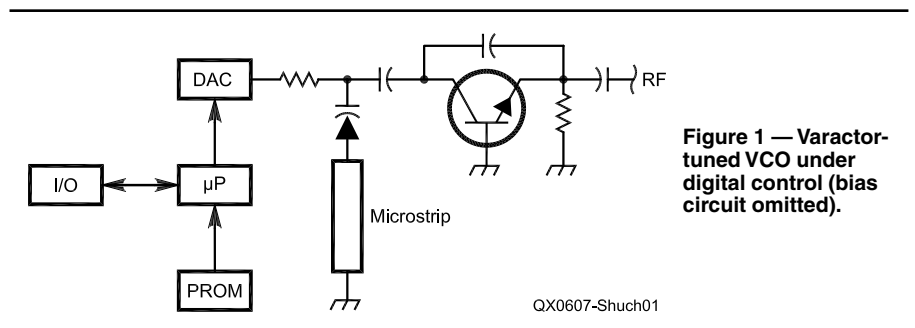
Dr H. Paul Shuch, N6TX

Introduction

Telecommunications systems are increasingly dependent on digital control of such tunable circuits as voltage-controlled oscillators (VCOs), electronically tuned phase shifters and attenuators, and filters incorporating ferro-tunable materials. The operating range of such circuits is typically limited by the linear dynamic range of their control components. A paradigm shift on the part of the designer can significantly increase the operating ranges of tunable circuit elements.

We've all worked with tunable circuits in our pursuit of amateur or professional electronic communications. Whether we drive such circuits directly with an analog tuning voltage, or indirectly with a digital word (which typically is converted to a tuning voltage in a digital-to-analog converter, or DAC), we notice that the linear dynamic tuning range of our devices is always less than ideal. So, we hope and cope; or we "complexify" and adopt closed-loop solutions, such as phase-locked loops, wherein we apply a control signal, sample its resulting circuit response, and use the observed response to modify the control signal, iterating that process until things settle down. Aside from the challenges of overshoot and undershoot that closed-loop solutions present, some kinds of applications are better served open-loop. It's often faster, cheaper and more reliable. So, can the linear tuning range be extended by compensating for nonlinearities? Might there be an open-loop solution to the problem?

I believe the answer is a resounding "Yes." Let's consider the example shown in Figure 1: a varactor-tuned VCO driven through a DAC by a microcontroller. The frequency response of this VCO is noticeably nonlinear

Vice President and Chief Technology Officer
QorTek, Inc
1965 Lycoming Creek Road, Suite 205
Williamsport, PA 17701
n6tx@arrl.net

Figure 1 — Varactor-tuned VCO under digital control (bias circuit omitted).

with tuning voltage, because the varactor's transfer function (*C* versus *V*) fits a perfect square-law over only a portion of its dynamic range, as seen in Figure 2.

Our objective is to control the oscillator in a predictable, repeatable and consistent manner. That means that a given control voltage needs to always produce the same output frequency. It's also desirable that each incremental change in the control signal produces a uniform frequency change. Conventional practice is to restrict the operating range of the oscillator to the circuit's linear tuning region. That achieves the desired consistency, predictability and repeatability, at the expense of limited performance.

Why Think Linear?

If you received your engineering education in the slide-rule era, as did I, you've probably characterized a tunable circuit or two. It's tedious work. You set a control voltage, read the resulting response, and write the two values in adjacent columns in your lab notebook. Then you set a new control voltage, take a new response, write the two, and repeat as necessary until you've filled the page. Next, it's time to plot response versus stimulus as dots on a piece of graph paper, lay a straightedge on the dots and visually determine a best-fit line through the dots. From that, it's easy to see the range of values over which the transfer function can be considered linear. (If we want to get fancy, we can extrapolate that
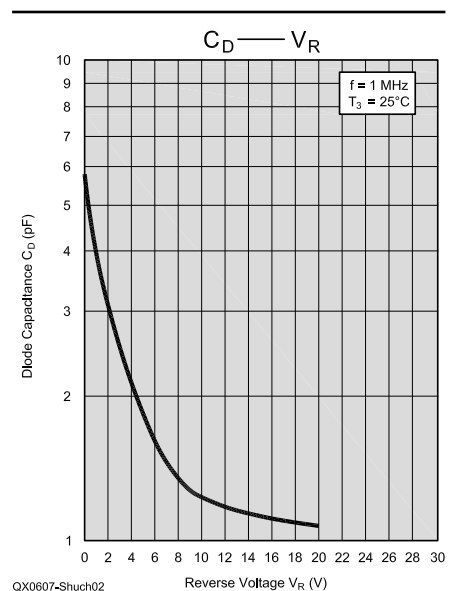
Figure 2 — C-V response curve for a typical silicon epitaxial varactor diode.

line until it intersects the axes, measure slope and intercept, and derive a linear equation in the form $y = mx + b$.) The extent to which the dots visibly diverge from the line is a rough indication of the quality of the fit, or correlation coefficient. Whether or not they told you in electrical engineering school, what you've just done is perform a visual linear regression analysis.

The process is a useful training exercise in that it not only provides us with a transfer function, it makes its derivation concrete and intuitive. As we shall see shortly, however, it also forces us to don the blinders of linear thinking — an obstacle that can be overcome only by making a concerted effort.

The Signpost Ahead: Entering the Nonlinear Zone

Once we've developed the transfer function as described above and reduced it to a linear equation, it's trivial for a computer to solve $y = mx + b$ "on the fly" for any desired circuit response. Of course, we need to program some hard limits into software so that we don't exceed the device's linear tuning range. If we tell the computer, in essence, "Neither go above 10 V nor below 2 V," we can maintain tuning linearity under digital control. By doing that, however, we've limited ourselves to only a portion of the performance that would otherwise be available to us.

There is a widely used work-around. By characterizing the oscillator over its full tuning range, we can develop a lookup table of the voltages required to achieve any number of output possibilities, whether linearly related or not. Said data table can be stored in programmable read-only memory (PROM) and accessed by the control computer, so that each twist of a tuning dial or click of a chopper wheel will produce a uniform response, even outside the circuit's linear tuning range. See Table 1. The process, which you can think of as linearization in software, extends the practical operating range and compensates the inherent nonlinearity of its response curve by stepping the control signal in a suitably nonlinear manner.

All this is very well and good if the particular operating point we desire corresponds exactly to a dot on our graph, or an entry in our lookup table. Generally, however, unless we have an infinite lookup table, we're forced to interpolate between data points. A linear interpolation between nonlinear points is bound to introduce a certain amount of inaccuracy. It may be close enough for government work, but is it close enough for Amateur Radio?

Going Polynomial

If only we knew the true underlying transfer function between data points, we could do a more accurate nonlinear interpolation. Instead of a linear transfer function, why didn't I simply derive a true polynomial function; say, in the form $y = ax^3 + bx^2 + cx + d$?

The answer is simply that I was educated in an era when such equations didn't simply pop out of the paper. Let's face it: We can do graphical nonlinear curve fitting with a combination of a compass, a French curve and a straight-edge; but visual derivation of the underlying equations is a skill not often taught within an EE curriculum. One useful trinket from analytic geometry is this: If you can draw a straight line that intersects a curve at n points, then the curve is at least an nth-order curve.

Modern digital computers are quite adept at curve-fitting. Throw a handful of data points at them, and they will return an elegant n^{th}-order equation that smoothly connects the dots. And if one computer can derive such an equation, another one can surely implement it.

So, I'm proposing that we banish linear thinking, dispense with the lookup table approach, and replace it with a nonlinear transfer function to be solved in software. All that remains is to determine the transfer function's coefficients.

Nonlinear Regression Without Aggression

Table 1 depicts the tuning range for a VCO similar to our Figure 1 sample, as entered into a Microsoft *Excel* spreadsheet. You will notice that by applying control voltages in the 0 to 25 V dc range, we can tune the oscillator's frequency across the 13-cm amateur band. It's hard to tell this from the raw numbers, but the relationship between voltage and frequency is nonlinear. We can nevertheless approximate it with a linear equation by performing the linear regression depicted graphically in Figure 3.

Table 1
Tuning response of a VCO

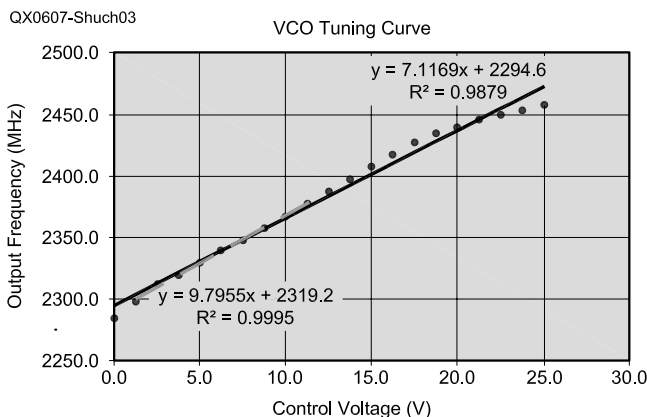| Bias (V dc)	Frequency (MHz)
0	2285.0
1.3	2297.5
2.5	2312.5
3.8	2320.0
5.0	2330.0
6.3	2340.0
7.5	2347.5
8.8	2357.5
10.0	2367.5
11.3	2377.5
12.5	2387.5
13.8	2397.5
15.0	2407.5
16.3	2417.5
17.5	2427.5
18.8	2435.0
20.0	2440.0
21.3	2446.3
22.5	2450.0
23.8	2453.8
25.0	2457.5

Figure 3 — VCO tuning curve with linear regression, over its full tuning range (solid line), and with the data set truncated to analyze only its most linear region (dashed line).
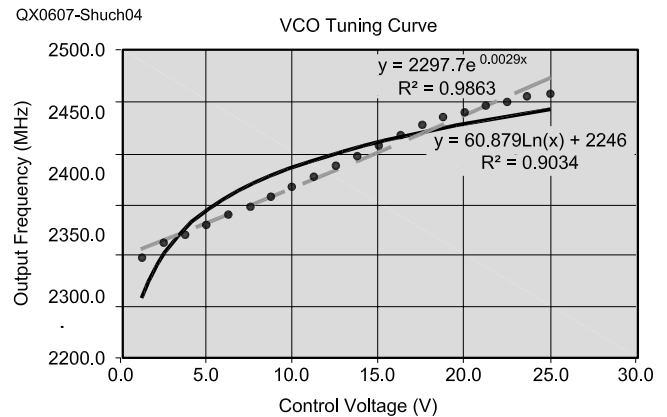

Figure 4 — VCO tuning curve showing logarithmic (the more curved line) and exponential (the more straight line) regressions.

The solid trend line in Figure 3 could easily have been established by employing the standard linear regression equations found in any basic statistics textbook. Fortunately, there is an easier way. Since we have already entered our raw data into a Microsoft *Excel* spreadsheet, why not let *Excel* plot a curve? Once that's done, we can employ the powerful curve fitting capabilities of that and other spreadsheet programs. Once the scatter diagram has been selected in *Excel*, you click on "Add Trendline" in the "Chart" pull-down menu. Options in the "Add Trendline" menu allow you to display not only the linear equation, but the correlation coefficient (R^2) as well. In this example, many of the data points miss the line but the correlation is a quite-respectable 0.9879 overall.

Can we do better? It appears from the graph that the response curve deviates from linear at higher tuning voltages. If linear tuning is required, we could restrict our operation to, say, 12 V or less. Truncating our data set in the spreadsheet and repeating the plot-and-regress routine, we generate the curve shown as a dashed line in Figure 3. Here, by restricting our tuning voltage range to 1 to 11 V dc (and our corresponding frequency range to 2330 to 2430 MHz), our data points all fall very nearly on the trend line, and the correlation coefficient goes up to a nearly perfect 0.9995.

Regrettably, we've achieved that significant improvement in linearity at the expense of frequency range. Instead of covering more than 170 MHz of S-band spectrum, we are restricting our tuning to a mere 100 MHz. Where only limited tuning is required, truncation is a viable technique for achieving tuning linearity. That's not necessarily the best approach, however.

Fortunately, in addition to linear regression, polynomial curve-fitting is one of the things at which spreadsheets *Excel*. Experimenting with different regression

options may give us a better fit over a wider range. So, let's experiment.

The arc in Figure 4 (let us ignore the more straight line there for the moment) shows the data for our VCO, fit to a logarithmic curve. It was necessary to omit the first data point corresponding to 0 V dc, since the logarithm of 0 is undefined.

Most of the data points miss the regression line by a country mile! The correlation coefficient R^2 is actually lower than it was for our full linear regression — clearly we're going the wrong way. Let's try something else.

How about a power curve? Again, we need to delete our first data point — this time because division by zero is not allowed. The solution is shown as the arc in Figure 5 (once again, ignore the straighter line for right now), which looks just as bad as the logarithmic regression. Back to the drawing board.

Using an exponential equation (turn back to Figure 4, this time concentrating on the straighter line) gives us a better fit, but it's still not quite as good as our initial linear approximation. Are we ever going to find a curve that fits our data?

Maybe we should try a quadratic: a second-order polynomial. *Excel* lets us fit to polynomial curves of arbitrary order, so two clicks produce the straighter of the two curves in Figure 5. For the first time, we see a relatively close fit between our curve and our data. The correlation coefficient R^2 is actually higher than it was for our first, linear regression (though still not as high as it was for our truncated linear data set).

Might this be a case where if a little is good, a lot is better? We can try higher-order polynomials. Figure 6 fits our data to a third-order curve, or cubic equation; for the first time, our full data set shows a correlation that exceeds even the linear fit for truncated data. We seem to have found a curve that fits our data well, and our spreadsheet has computed the transfer function for us.

Why quit when we're ahead? Fourth, fifth, and sixth-order polynomials were tried. In each case, the correlation coefficient comes that much closer to unity — a perfect fit. Remember, though, that at some point, these curves are going to have to be solved in software. The higher the order of the polynomial, the more memory, time, and computer cycles will be consumed in the calculation. It appears from my tests that we're rapidly starting to see less bang for our buck. Although the sixth-order polynomial shows a nearly perfect fit, with $R^2 = 0.9998$, that's no better than the results from the fifth-order curve. So we've actually gone "beyond the point of no return."[2]

Table 2 summarizes our results thus far. The third-order polynomial seems to be our best bet. Beyond that, the regression coefficients for the higher-order terms can be seen to rapidly be approaching zero. As will be seen in the case studies that follow, it happens that the third-order polynomial is generally the curve of choice for a wide variety of voltage-tuned electronic circuits.

As it happens, the variation in R^2 values seen in Table 2 is not all that dramatic, because this particular circuit had pretty good tuning linearity to begin with. Such is not always the case, however, and in other less-linear circuits, the improvement can be even more impressive.

Case Study 1:
Phase Shifters for the Very Small Array

The VSA is a technical initiative of the membership-supported, nonprofit SETI League to develop the next generation of research-grade radio telescopes by arraying a large number of backyard satellite TV dishes.[3] Because the signals from all the dishes must be summed in phase to maximize gain and angular resolution, and since variations in group delay between even seemingly identical components are inevitable, it became necessary to fabricate a network of

Table 2
Comparison of Correlation Coefficients for various models

Curve	R^2
Linear	0.9879
Truncated Linear	0.9995
Logarithmic	0.9034
Power Curve	0.9079
Exponential	0.9863
Second Order Polynomial	0.9965
Third Order Polynomial	0.9987
Fourth Order Polynomial	0.9994
Fifth Order Polynomial	0.9998
Sixth Order Polynomial	0.9998

QX0607-Shuch05

VCO Tuning Curve

$y = -0.1027x^2 + 9.6917x + 2284.3$
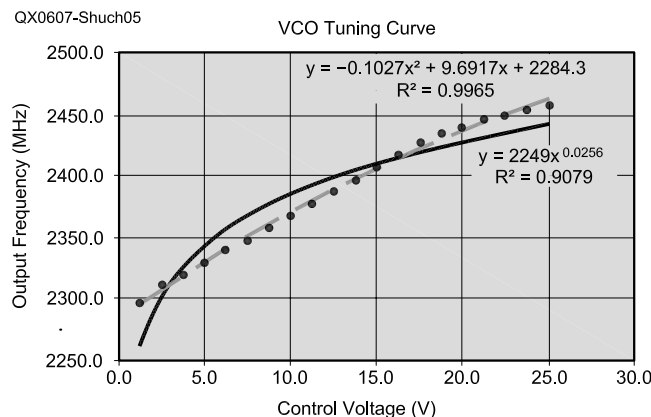$R^2 = 0.9965$

$y = 2249x^{0.0256}$
$R^2 = 0.9079$

Figure 5 — VCO tuning curve showing power (the more curved line) and 2nd-order polynomial (the more straight line) regressions.

voltage-tuned analog phase shifters. Those are adjusted *in situ* to compensate for phase inconsistencies and thus achieve optimal system performance.

L-band phase shifters designed for the VSA offer 0 to 90° of phase shift across the lower microwave spectrum, with modest linearity, when tuned with a 0 to 10 V dc control voltage. Central to the operation of this design is a precision etched microstrip-line quadrature hybrid coupler, which produces equal-amplitude in-phase (I) and quadrature (Q) components of the applied signal. The quad hybrid is terminated in a pair of matched varactors. Reverse-biasing the varactors varies their junction capaci-tance, and the resulting variable reactance termination on the hybrid causes the input signal to be reflected back to the output terminal with varying phase. The high linearity of the unit's phase response is achieved by limiting the applied bias to the linear portion of the varactors' C/V curve.

If compensation for varactor nonlinearity could be applied, it should be possible to increase phase tuning beyond the specified 90° range. The tuning linearity of one such phase shifter is depicted as the straight line in Figure 7. Note that the curve is nonlinear at its extremities.

If required, we can truncate the curve to achieve nearly ideal linearity. The resulting operating region, however, is reduced to about 60° of linear tuning. By fitting a third-order polynomial curve to the observed performance data for this phase shifter, as seen in the curved line of Figure 7, the correlation coefficient increases from 0.9639 to a nearly perfect value of 0.9994, over a 95° tuning range.

Case Study 2:
Millimeter-Wave Microbolometers

In my day job, under contract from the US Air Force Research Laboratory, I recently characterized some chromium-on-polyimide bolometers fabricated by my colleague Professor David Nelson and his graduate students at Michigan Technological University.[4] Because these microwave power-measuring devices are strongly influenced by ambient temperature, there is quite a bit of noise in the measured power response. Their response fits well to a third-order polynomial curve, however, with R^2 increasing to 0.9974, as compared to 0.9680 for the linear model. The two curves are compared in Figure 8.

Case Study 3:
Ferro-Tunable Antenna Elements

Under contract to the NASA Earth-Sun Science Technology Office at Goddard Space Flight Center, I've been privileged to serve as Principal Investigator on the development of a new space-borne synthetic-aperture radar (SAR) antenna.[5] This antenna is unique in that it employs a ferro-tunable substrate of
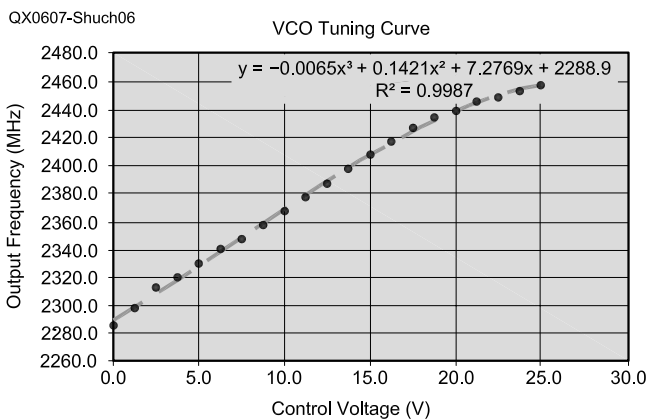

QX0607-Shuch06

VCO Tuning Curve

$$y = -0.0065x^3 + 0.1421x^2 + 7.2769x + 2288.9$$
$$R^2 = 0.9987$$

Figure 6 — VCO tuning curve with 3rd-order polynomial regression.

QX0607-Shuch07

L-Band Phase Shifter

$$y = 9.7169x + 10.892$$
$$R^2 = 0.9639$$

$$y = -0.032x^3 - 0.2052x^2 + 14.698x + 1.404$$
$$R^2 = 0.9994$$
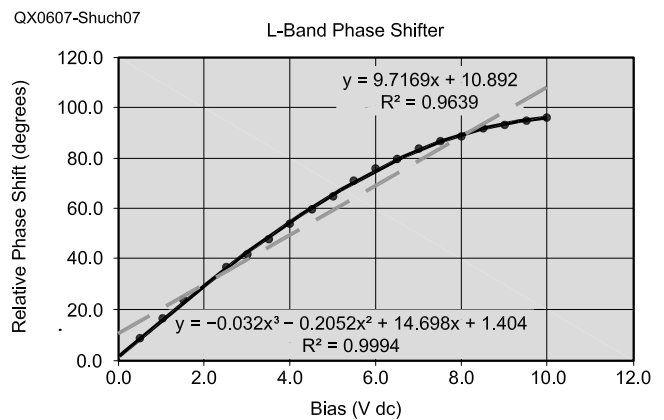
Figure 7 — Tuning curve of a phase shifter, showing linear (the straight line) and 3nd-order polynomial (the curved line) regressions.

QX0607-Shuch08

Thin-Film Chromium Bolometer

$$y = 0.006x + 46.742$$
$$R^2 = 0.968$$

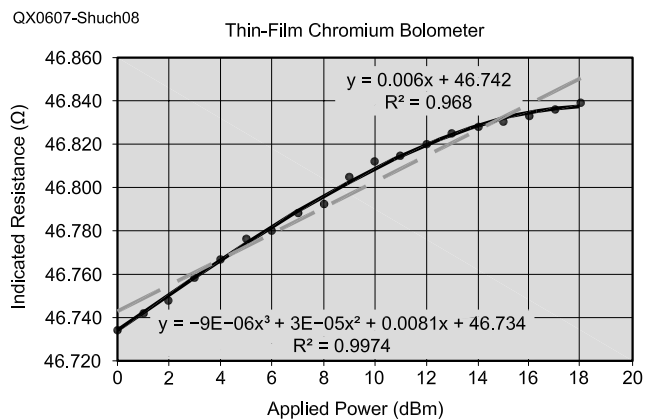$$y = -9E-06x^3 + 3E-05x^2 + 0.0081x + 46.734$$
$$R^2 = 0.9974$$

Figure 8 — Transfer function for a thin-film chromium bolometer. The dashed straight line is a linear model. The solid curved line represents a third-order polynomial model.
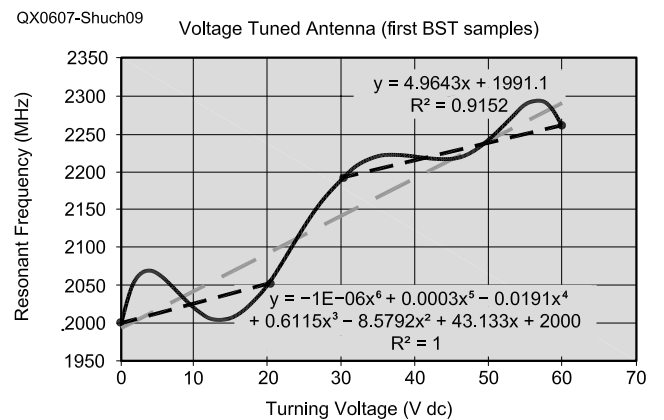

QX0607-Shuch09

Voltage Tuned Antenna (first BST samples)

$$y = 4.9643x + 1991.1$$
$$R^2 = 0.9152$$

$$y = -1E-06x^6 + 0.0003x^5 - 0.0191x^4$$
$$+ 0.6115x^3 - 8.5792x^2 + 43.133x + 2000$$
$$R^2 = 1$$

Figure 9 — BST antenna tuning curve. The light gray dashed line represents linear regression. The curved line represents 6nd-order polynomial regression. The two black dashed lines probably represent reality (see text).

barium-strontium-titanate (BST) developed by my colleague Professor Jon-Paul Maria and his graduate students at North Carolina State University. The tunable substrate allows us to vary the antenna array's resonant frequency by application of dc voltages. BST isn't a highly linear material, as evidenced by the deviation of the measured data points form the solid straight line in Figure 9. See Figures 10, 11 and 12 for photographs of the array. Once the non-linear tuning algorithm for the steerable and tunable array has been validated, it is implemented in software. Figure 13 is a screen shot of the software display.

Sixth-order polynomial regression produces a matching curve with a perfect fit ($R^2 = 1$), as seen in the curved line in Figure 9. A word of caution is in order. Just because we can find a function that fits the available data, we shouldn't assume that function reflects reality.

It's a mathematical truism that *any n* data points will *perfectly* fit an $(n-1)^{th}$ order polynomial equation. In the present case, with only seven data points available, it should come as no surprise that our correlation with a sixth-order curve is ideal. Prudence and the sample-size rule dictate that the number of data points being evaluated must always significantly exceed the order of the polynomial

Computational Algorithms

Designers of embedded computer systems face trade-offs between computational accuracy, complexity and speed. Computing integral powers of numbers in polynomial functions is straightforward. Nonintegral powers and transcendental functions generally require algorithms: processes with more than one step.[7]

Lookup Tables

Lookup tables constitute the fastest way to find values of almost any function but they tend to occupy large amounts of memory. The processes of decimation and interpolation achieve smaller tables at the expense of accuracy. Decimate a large table by a factor of N by keeping only every N^{th} entry. Interpolate during lookup by finding the two table addresses closest to the argument and making an estimate of the target value between the two table values. A first-order estimate between two values is linear: a straight line. That results in a linear piecewise representation of the function.

An address into a decimated table is truncated. It doesn't retain the full precision of the original argument. Looking up two values from a decimated table, based on the full precision of the argument, always produces a single result of greatly improved accuracy. How the two addresses are formed depends on the nature of the function. Square-roots and base-two logarithms are my examples.

Let's say we need to find the square root of the 16-bit binary number M = 0110 1001 0111 1000$_2$ = 27,000$_{10}$. A lookup table of 2^{16} entries is out of the question. Begin by constructing a table decimated by N = 2^8 (with 2^8 entries). Divide the argument by 2^8 by keeping only the 8 most-significant bits: N = 0110 1001$_2$ = 105$_{10}$. Now look up the square root of N from the table and the initial answer is: M$^{1/2}$ ≈" (2^8)$^{1/2}$ N$^{1/2}$ ≈" (2^4)(105)$^{1/2}$ ≈" 163.951 (error 0.2%). Obtain a second-order estimate using the relation M$^{1/2}$ ≈" 2^4 N$^{1/2}$ / [(2^8 N) / M]$^{1/2}$, where only 8 bits of the quotient (2^8 N) / M need be retained as the second lookup address (net error 0.03%).

The same kind of algorithm works for base-two logarithms. Again let M = 0110 1001 0111 1000$_2$ = 27,000$_{10}$ and N = 0110 1001$_2$ = 105$_{10}$. Now log$_2$ (2^8 N) = 8 + log$_2$ N is close to the correct answer and log$_2$ N = 6.714246 may be looked up in a 256-entry table (error 0.04%). A closer estimate may be computed using the relation log$_2$ M = 8 + log$_2$ N − log$_2$ (2^8 N / M) where only 8 bits of the quotient are retained. Those 8 bits are convergently rounded and again form a second address into a 256-entry lookup table. The result, 14.7199, comes out a little low (error 0.005%).

Taylor Series

Taylor's formula is a useful tool in curve-fitting. It effectively produces a polynomial approximation for any continuous, differentiable function, and a remainder (error). Computation of the first few terms of Taylor series for transcendental functions is a popular embedded processing method that's slightly slower than lookup tables but much more efficient in its use of memory. For example:

$$\sin x = x - x^3 / 3! + x^5 / 5! - x^7 / 7! + \ldots$$

Those four terms produce a result with a maximum error ($0 \leq x \leq \pi/2$) of about 0.0079%.

Iterative Methods

Iterative methods are also popular in embedded systems. An iterative method produces an output at each iteration that becomes the input for the next iteration. Computing the roots of a differentiable function is often done with what's called Newton's method. To go from the nth estimate x_n to the $(n+1)^{th}$ estimate x_{n+1}, use the equation: $x_{n+1} = x_n + f(x_n) / f'(x_n)$, where $f'(x_n)$ is the first derivative of f at x_n.

For a square-root calculation, we're solving the equation $x^2 - y = 0$ for x, where y is the argument. With $f(x_n) = x^2$ and $f'(x_n) = 2x$, the governing equation reduces to $x_{n+1} = (x_n / 2) + (1 / x_n)$, which converges to a result accurate to within 0.001% in five or six steps, depending on the initial guess. Note that two roots exist in this case: one is positive and the other, negative.

Newton's method doesn't always converge. Even when it does converge, it may not converge on the expected root where multiple roots exist. A good initial guess is important because in certain cases, the outcome may be extremely sensitive to the starting value. The function $f(x) = 4x^4 - 4x^2 = 0$ is a case in point. Three roots exist (−1, 0 and +1) and initial guesses within certain ranges produce an unexpected solution or do not produce a solution. Advanced calculus provides tests for convergence that must be used to avoid ambiguity and failure.

Picard's method is an iterative method that solves equations of the form $f(x) = x$. Any root-finding equation can be placed in that form simply by adding x to both sides. To find the square root of 0.5, for example, $f(x) = x^2 - 0.5 = 0$ becomes $g(x) = x^2 - 0.5 + x = x$. Now the iterative equation is $x_{n+1} = x_n^2 - 0.5 + x_n$. The method diverges when the initial guess falls where $|g'(x)| \geq |1$. In such cases, a solution for $g^{-1}(x)$ is possible instead.

Other Algorithms

Mathematicians have discovered many more advanced root-finding algorithms. Speed of convergence on quadratic and higher-order performance surfaces is the subject of ongoing investigation. Other communications applications of such algorithms include noise reduction, adaptive beamforming and channel equalization.
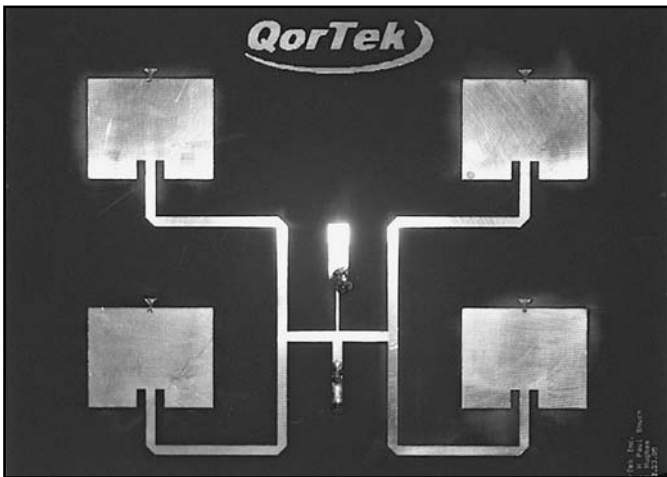— *Doug Smith, KF6DX,* QEX *Editor.*

Figure 10— This photo is a close-up view of a steerable and tunable synthetic aperture array.
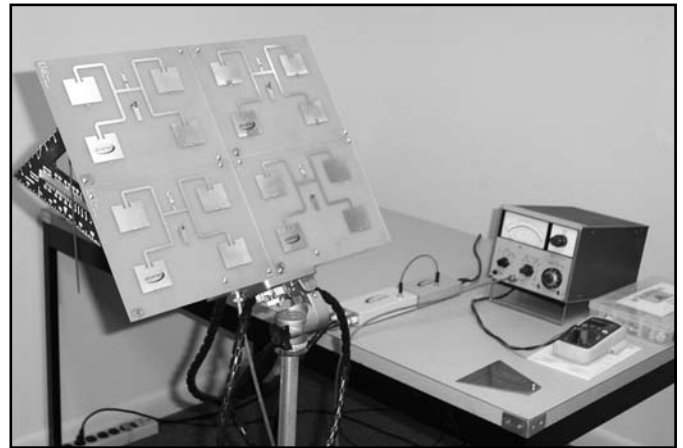

Figure 11 — A steerable and tunable synthetic aperture array being tested in the author's laboratory, to quantify its tuning transfer function for nonlinear analysis.

being used to characterize them.

Although this project (depicted more fully in Figures 10 through 13) is still a work in progress, I'm prepared to speculate as to what's really going on. Microstrip patch antennas (such as are implemented here on a ferro-tunable dielectric substrate) can be described as cavity/slot radiators. Anyone who works with microwave resonant cavities can tell you that they can operate in multiple modes. I conclude that our antenna is jumping between two distinct modes, each of which exhibits a relatively linear frequency response. For this circuit, two disjointed linear curves (shown as dashed lines in Figure 9) fit the data well. So does a sixth-order polynomial. Mathematically, either solution yields a satisfactory result. But, as far as antenna tuning is concerned, which interpretation do you think is more likely to represent reality?

William of Ockham said it best.[6] When presented with two potential explanations for a phenomenon, the simpler of the two is most often correct. Graphing circuit responses enables us to apply the Inter-Ocular Trauma Test. We accept as true those relationships that are strong enough to strike us between the eyes!

Pitfalls to Avoid, Data Points to Ponder

You may be wondering why I wrote that *any n* data points will perfectly fit an $(n-1)^{th}$-order polynomial. Consider that an nth-order polynomial actually has $(n+1)$ terms, with exponents of n through zero. That lowest-order term is just a numerical coefficient, since $x^0 = 1$ for any x. So, if we fit an $(n-1)^{th}$ order polynomial to any arbitrary set of n data, the resulting curve will go precisely through all the points. At this juncture, our polynomial is not really an equation but rather a lookup table expressed in
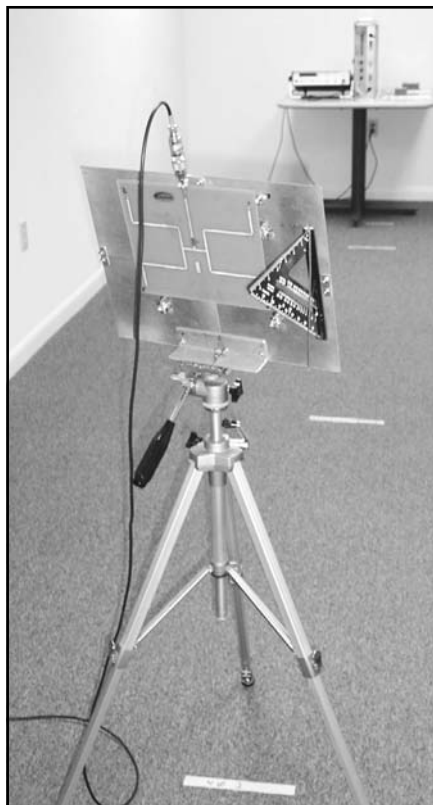

Figure 12 — The NASA synthetic aperture array being tested on a calibrated compact antenna range. This test will verify proper tuning and steering control, based on the nonlinear transfer function determined on the test bench.

polynomial form. Wouldn't you expect all the points that go into a lookup table to correlate exactly with that lookup table? It should be no surprise that very high-order polynomial curve-fitting produces results that are completely accurate but may be useless between data points.

I stated earlier that any polynomial equation that one computer can derive, another can implement. Given the appropriate software, one can solve for the independent variable (typically, tuning voltage) necessary to drive any circuit, producing the desired dependent variable (frequency, gain, attenuation, phase shift, whatever). One must not overlook the speed at which such computations occur, however. There is a limit to how quickly one can solve a third-order polynomial in software.

High-level computer languages and graphical user interfaces, popular though they are, may impose unacceptable overhead in such control applications, suggesting that it may be advisable to implement the tuning computations in machine code or assembler. The level of computational precision must also be balanced against available memory and processor speed. In some applications, the rate at which a circuit's drive signal must update can readily exceed the required computational time. Thus, a simpler model may be required, even at the expense of reduced precision. If we're converting a digital word to a control signal in a DAC, some thought must be given to the number of bits and level of resolution actually required, as this too impacts processing time. See the sidebar "Computational Algorithms" for additional information.

Conclusions

It's easy to get locked into linear thinking, viewing any nonlinearity in circuit performance as a design deficit. This conditioned response to growing up in a graph-paper-and-straight-edge world is a pitfall that can limit our vision. By recognizing that high-order polynomials are readily solved and easily implemented in software, and that a circuit's transfer

function is rarely, if ever, linear, we can undergo a paradigm shift that enables simplicity, reliability, increased tuning range and improved circuit performance.

Notes

[1]Tim Rice (lyricist) and Andrew Lloyd Webber (composer), "The Point of No Return," from *Phantom of the Opera*.

[2]H. Paul Shuch, "The Very Small Array," *QST*, Sep 2002, pp 28-30.

[3]H. Paul Shuch, "Large Area Millimeter Wave Dosimetry," Final Report, Contract FA8650-05-M-6506 Topic No. OSD04-H05, Air Force Research Laboratory, Aug 5, 2005.

[4]H. Paul Shuch, "Multiband Reconfigurable Synthetic Aperture Radar Antenna." *Proceedings of the 2004 Earth Science Technology Conference: B1P1*, NASA, June 2004.

[5]H. Paul Shuch, "Updating the Multiband Reconfigurable Synthetic Aperture Radar Antenna," *Proceedings of the 2005 Earth-Sun System Technology Conference: B5P1*, NASA, June 2005.

[6]The 14th Century English philosopher and theologian best known for his medieval rule of parsimony, which came to be known as **Ockham's razor**: *plurality should not be assumed without necessity.*

[7]G. Thomas, Jr. and R. Finney, *Calculus and Analytic Geometry,* Addison-Wesley, Reading, MA, 1993, ISBN: 0-201-52929-7.
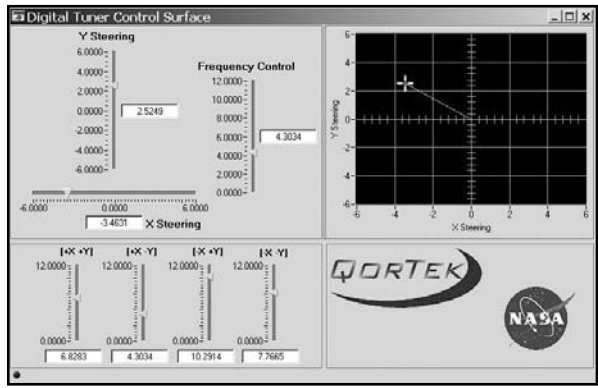
Figure 13 — Once the nonlinear tuning algorithm for the steerable and tunable array has been validated, it is implemented in software. This graphical user interface allows full array control (azimuth, elevation and frequency) from a laptop computer, with drag-and-drop input.

About the Author

H. Paul Shuch, N6TX, enjoyed perhaps the shortest retirement on record, leaving academia on a Friday in 2004, and being put on payroll by a consulting client the following Monday. Now Vice President and Chief Technology Officer of QorTek, Inc, a small start-up Pennsylvania government contractor, Paul is responsible for several NASA and DoD engineering and research projects dealing with a wide variety of electromagnetic systems.

A distinguished engineering professor, Dr. Shuch earned his Ph.D. in Engineering from the University of California, Berkeley, after serving in the US Air Force as a telecommunications systems controller during the Vietnam conflict. Founder and chief engineer of Microcomm, the Silicon Valley startup credited with developing the world's first commercial home satellite TV receiver, he served as Technical Director and Chairman of the Board of Project OSCAR Inc, builders of the first non-government communications satellites.

While an engineer with such major aerospace corporations as Itek and Lockheed, Dr. Shuch was responsible for the design of electronic countermeasures, satellite remote sensing, and submarine launched ballistic missile telemetry systems. He holds patents in the areas of airborne radar and phased array antennas, and taught electronics and avionics at various colleges and universities for a quarter century. Dr. Shuch is the author of over 400 publications, many of which have graced the Amateur Radio literature. Paul is a member of the International Academy of Astronautics and the AACS Alumni Association; vice president of the Society of Amateur Radio Astronomers; a life member of the Society of Wild Weasels, AMSAT, and ARRL; a Fellow of the British Interplanetary Society and the Radio Club of America; a member of the AMSAT, SARA, and SETI League boards, and serves on numerous international committees. **QEX**

Effective Directivity for Shortwave Reception by DSP

Learn how the author uses two small loop antennas and synchronized receivers with DSP software to reduce or eliminate interfering signals.

Jan Simons, PAØSIM

R adio amateurs living in a suburban or residential location are faced with some limitations. The most important limitations are the space for placing shortwave antennas, a high ambient man-made-noise level and the significant possibility of local interference. On the top bands, large directional antennas are hardly an option. Given these preconditions, we can optimize shortwave reception by balancing antennas as well as possible, and by placing them as far as possible from local interfering sources. In some cases, a local interfering source can be attenuated by using the null in the radiation pattern of a small "magnetic" loop antenna. With the help of noise cancelling and a second antenna, a single local interfering source can always be suppressed. Not much more is practically feasible with analogue means in a residential location, however.[1]

How can digital signal processing (DSP) help to improve shortwave reception, whereby we do not wish to limit ourselves to one or more local interfering sources? It turns out that DSP, using time difference of arrival (TDOA), in combination with two active small loop antennas and two synchronized receivers, is capable of creating an effective directivity, which is comparable to large directional antennas.

Frequency Domain Processing

As an example, we will take a close look at the content of an SSB signal, as shown in

Prins Mauritsstraat 14
Venlo, 5923 AZ
The Netherlands
pa0sim@amsat.org

Figure 1. The bandwidth needed for a voice channel is around 3 kHz. During every fraction of a second however, only a part of that channel is used. There are always only a limited number of frequency components present.

Depending on the speech content in the signal, the channel will contain different frequency components every following moment. Every interfering source, whether local or not, will also contain its own frequency components at any moment. The chance that frequency components of different sources will overlap one another is very small. See Figure 2.
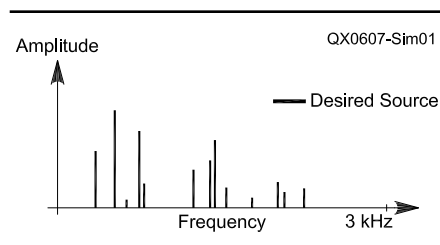

Figure 1 — Limited number of frequency components per unit of time of a single source.
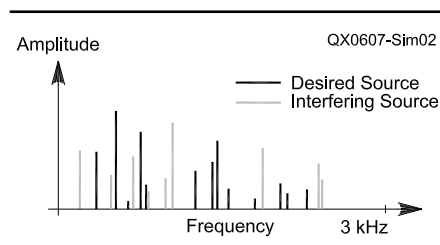

Figure 2 — Non-overlapping frequency components from two sources.

If we are able to select only the desired frequency components, this is comparable with a reduction in bandwidth. Those parts of the spectrum in which we are not interested are cut off — just as with a CW filter, but within the channel. This assumes that each frequency component always comes from just one source. This is of course, not entirely true, as there is noise from the antenna preamplifiers and atmospheric noise over the whole band in variable amplitude. There is always a small chance that more than one source will contribute to a frequency component. We will deal with this later in more detail.

Distinguishing Frequency Components Using TDOA

How can we distinguish the desired frequency components from the interfering frequency components for the purpose of selection in the frequency domain? How do the various signal sources differ?

We can look for a difference in the typical characteristics per type of signal source. CW, SSB and ambient noise have different characteristics. The assumptions one must make for this do not always hold true. An interfering source can, for example, have the characteristics of speech or can even be speech.

A practical and more useable difference is seen in the direction (azimuth and elevation) from which the signals arrive. Direction translates, when using two or more antennas, into time difference of arrival (TDOA) or phase differences and in ratios of amplitudes.

We see the same effective difference whether or not signals are circularly polarized, and the sense of this circular

polarization. We will also see this as a phase difference later on.

We can now choose what we are going to do with each frequency component based on phase difference and/or amplitude ratio. We limit ourselves to discrimination based on phase difference alone. There are two reasons for this. First, because the phase difference mainly provides the information on the direction, and second, because the processing is carried out at the audio, and not at the intermediate frequency level. As a consequence of the AGC, it is not possible to determine the real amplitude ratio. It is mainly due to the AGC that this sort of processing must actually be carried out at the intermediate frequency level.

From Direction and Polarization to Phase Difference

We need a minimum of two antennas to achieve a directional phase difference. Two antennas that are small with respect to the wavelength are a good choice. See Figure 3.

For example, active small "magnetic" loop antennas work well (see Note 1). They are wideband, optimally balanced and there is a negligible mutual coupling between both antennas.

Loop Antennas in the Same Direction

It is obvious we are about to use both loop antennas as an array, as we are mainly interested in the direction from which the signals enter. In this array, all antennas must have identical radiation patterns and the polarization must also be the same in all
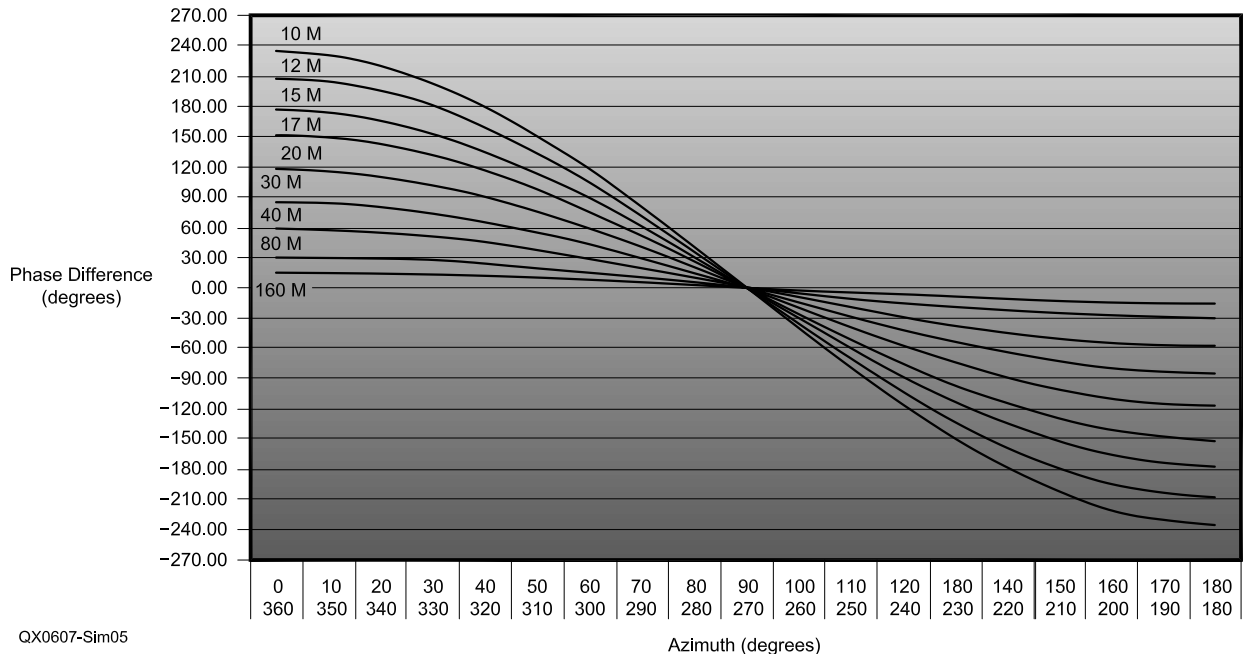

Figure 3 — Three configurations with the antennas a) in the same direction b) at right angles to one another and c) above one another.
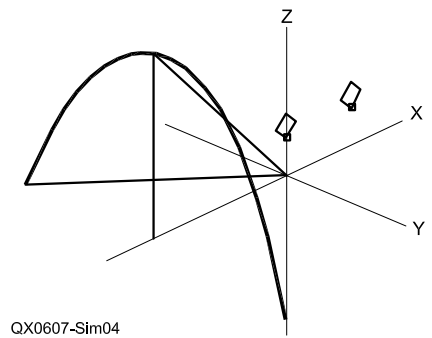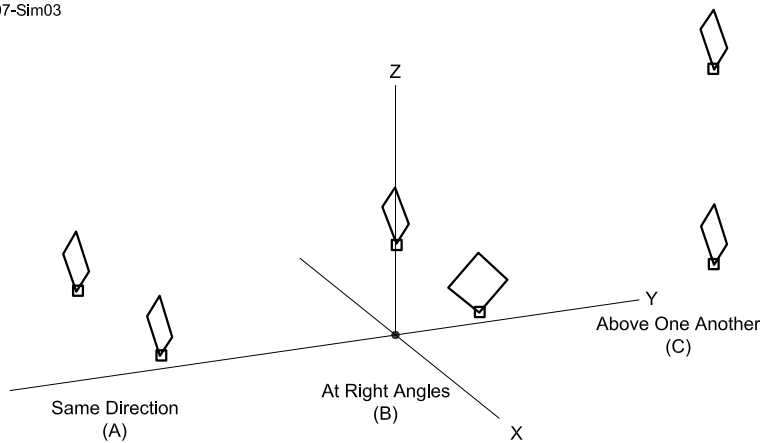
Figure 4 — Two antennas show an arc of directions with the same difference in phase.

Figure 5 — The phase difference as a function of the azimuth direction for a 0° elevation angle.

directions. The resulting radiation pattern of an array is determined by adding the fields of the antennas with suitable amplitude ratios and phase differences.

In our case, however, we are not going to add the fields, but decompose one field into two components (signals) with their own amplitude and phase. In this specific case, we look at the difference in phase between the signals received by both antennas. This actually concerns time difference of arrival (delay) differences. One antenna receives the signal earlier in time than the other antenna. We may convert this into a phase, as the bandwidth of the channel is small with respect to the carrier frequency.

This phase difference is simple to calculate. We already know the wavelength of the carrier frequency. Now we only have to work out the difference in distance between both antennas for the respective direction. A wavelength corresponds to 360° (a complete sine period). The phase difference is therefore this 360° times the ratio of the difference in distance and the wavelength. Depending on which antenna first receives the signal, this phase difference will be positive or negative.

Remember, the direction also includes elevation as well as azimuth. We get the same phase difference for an arc of direction, as shown in Figure 4.

Each arc is formed by all directions having the same angle with the axis on which both antennas are situated (X-axis). We can therefore make no distinction between signals on the opposite side of this axis. We need more antennas for this. Only in the extension of this axis and at 0° elevation do we find a single point where the phase difference is at a maximum. Figure 5 shows the phase difference as a function of the azimuth direction for an elevation angle of 0° and an antenna distance of 7 m for each band. With increasing elevation angle, these phase differences become smaller. Apart from the phase difference, we therefore also need the elevation angle in order to determine the azimuth direction. We cannot determine this elevation angle with two antennas, and can only estimate this based on the distance to the transmitter.

Loop Antennas at Right Angles to One Another

Here also, we are still looking only at the far field. When both antennas are at a relatively short distance from one another, a wave will arrive almost simultaneously. Ground waves and local interfering sources are linearly polarized. With small antennas, this can only result in a phase difference of 0° or 180°. I will elaborate on this in the "Near Field" section.

Radio amateurs normally do not dwell on it, but sky waves can be circularly (mostly elliptical) polarized. At 160 and 80 meters, this circular polarization can be highly constant and predictable. This leads to phase differences of about –90° or +90°. On the higher-frequency bands, polarization is normally linear and variable in direction. The phase difference wanders unpredictably over the entire 360°.

We can make use of this circular polarization. At 160 and 80 meters, there are now unexpected possibilities, as it is difficult on a residential location for those bands to have sufficient distance between both loop antennas. With the loop antennas in the same direction, the maximum phase difference on those bands is then quite small. In the case of local stations, this is further reduced by the high elevation angle (NVIS). The smaller the phase difference, the more difficult it is to separate the signals. Thanks to circular polarization, and with the loop antennas at right angles to one another, we can still separate local interfering sources and ground-wave signals from desired sky-wave signals.

Loop Antennas Above One Another

It would be a fine thing if we could select the signals based on the elevation angle. Local interference arrives at a very low elevation angle. The same, however, also applies to the true DX signals. Many other signals enter under a steeper elevation angle. These can then be separated from local interference.

This is simple in free space. You place the two antennas at a distance above one another. Depending on the elevation angle, we measure a phase difference.

In practice, the (loop) antennas are gen-erally, with respect to the wavelength l, placed on a low height (< ¼ l) above the ground. The elevation angle cannot be easily determined with two antennas in this condition.

Figure 6 shows the contribution made by ground reflection. We find this picture in Chapter 3 of *The ARRL Antenna Book* and other Amateur Radio handbooks.[2] The wave reflected by the ground appears to have been received or transmitted by a mirror antenna in the ground. Let us assume the ground reflects 100%. Then the mirror antenna receives the same signal strength, but at a distance BC, later. The real antenna receives the total of both signals, the direct and the reflected wave. The net phase is then precisely between the phases of the real and the mirror antenna. This is exactly the phase that we receive at a height of 0 meters (half the distance A – B between the real and the mirror antenna).

The phase we measure is therefore independent of the height at which we place the antenna. In reality, the ground is not ideal and we will measure phase differences. But these will be of no use until we have placed one antenna at a height of ¼ λ or higher. On a residential location, neighbors may find as little as 10 meters to be too high. This option will therefore be mainly applicable for the higher-frequency bands, 30 meters and higher.

Near Field and the Phase Difference

Up until now, we have mainly dealt with the far field. In a residential location, however, we can, for local interfering sources, easily be dealing with the near field. For the boundary between far and near field, a distance of 1/6 λ may be maintained, but this
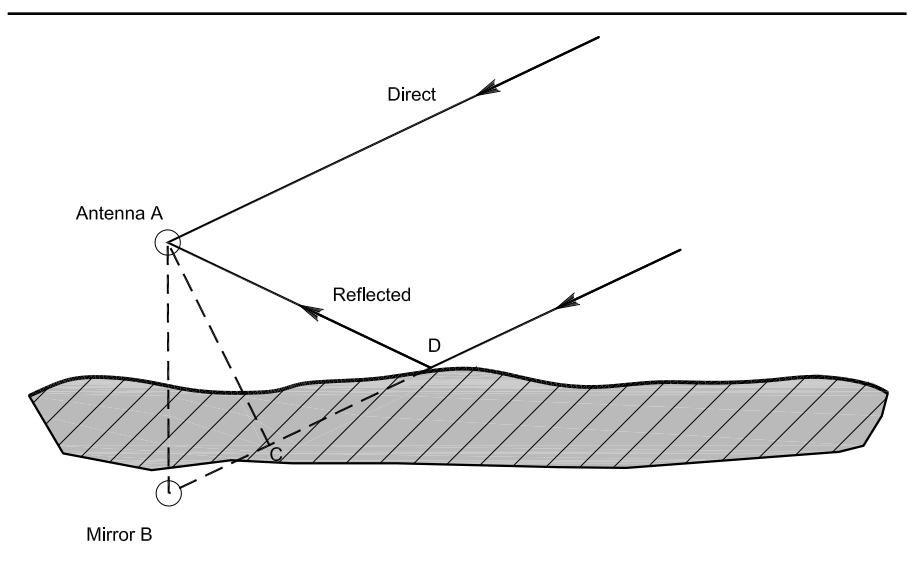

Figure 6 — Phase with low level antennas independent of height as a result of far field ground reflection.

is not a strict limit and must sometimes be set at 1 λ or more. The dimensions from the source with respect to the wavelength are very relevant in this matter. The near field radiation is completely different from that of the far field and the relationship between phase difference and direction is no longer obvious. Unexpected phase differences can occur, which, however, remain constant in time!

With the loop antennas at right angles to one another, this translates into a phase difference that is no longer in the 0° or 180° range. The phase difference can have any value for a large-sized local source. Overlap in phase difference of a local interfering source and a desired source can then occur by chance.

If we set the loop antennas parallel in the same direction, we can see something similar. The phase difference can be larger than we expect based on the distance between the antennas. This is favourable, as the local interfering source and a desired source cannot overlap in this case.

In this context, it should also be mentioned that coupling with objects (such as antennas) in the near vicinity can influence the phase difference. This coupling can normally be minimized by turning the loop antenna(s).

Overlap in Phase Difference

There is a chance that an interfering source gives the same phase difference as a desired source. Both sources can then no longer be separated based on phase difference. This phase difference applies with loop antennas in the same direction in the far field for an arc of directions and the chance of this is thus higher than with a very large directional antenna (beam). See Figure 4. By giving another azimuth direction to the axis on which both antennas are situated, we also give this arc another direction. This can solve overlap in phase difference. A practical solution to this is a third antenna on another axis direction. With interfering sources of relatively large dimensions with respect to the wavelength in the near field, all unexpected phase differences can occur. Turning both loop antennas can then sometimes solve overlap.

Elliptical Polarization as a Result of Propagation

Elliptical or circular polarization can be found in ionograms. Figure 7 shows an ionogram of the Belgian Dourbes.[3] The signals reflected by the ionosphere are shown in two shadings.

We can imagine a linear polarized field, such as that of a dipole for example, as consisting of the sum of a right-hand and a left-hand circularly polarized component. Together, they form the linearly polarized field. In the ionograms, we see both circularly polarized components as the ordinary and extraordinary wave. If now, due to propagation by the ionosphere, for example, the left-hand circular component is attenuated, a right-hand circular field remains.

On the top bands, the extraordinary wave (light) normally seems to have a higher attenuation. In these cases, the ordinary wave (dark) is dominant. We will then be left with a circular (normally elliptical) polarized field. This effect in propagation is of course not constant and depends on propagation, on the time of day, the time of year and solar activity.

There is, however, not much information about this in normal handbooks or on the Internet. On the 80-meter band, the ordinary wave appears to dominate during the day (phase difference around –90°) and the signals are normally circularly polarized. As it gets dark, this moves to linear polarization with a once dominating extraordinary wave (phase difference around +90°). On the 160-meter band, we at least see circular po-

STATION	YYYY	DAY	DDD	HHMM	P1	FFS	S	AXN	PPS	IGA	PS
Dourbes	2005	Mar30	089	1401	MMM	400-1	8c5	100	+1+	A2

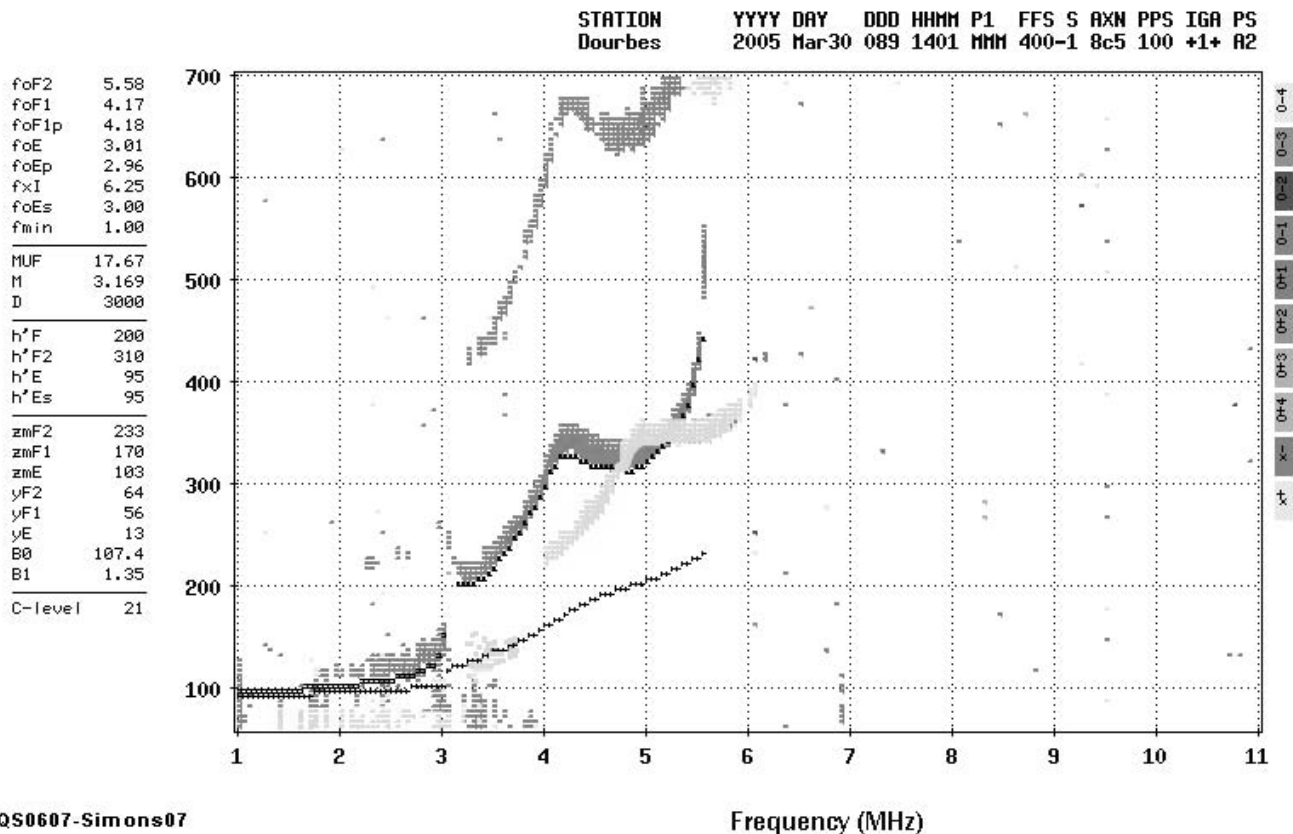|	
foF2	5.58
foF1	4.17
foF1p	4.18
foE	3.01
foEp	2.96
fxI	6.25
foEs	3.00
fmin	1.00
MUF	17.67
M	3.169
D	3000
h'F	200
h'F2	310
h'E	95
h'Es	95
zmF2	233
zmF1	170
zmE	103
yF2	64
yF1	56
yE	13
B0	107.4
B1	1.35
C-level	21

QS0607-Simons07

Frequency (MHz)

Figure 7 — Dourbes' ionogram with ordinary (light) and extraordinary (dark) reflections.

larization in the evening. Ground wave propagation is mainly vertically polarized, not circular. I have found no information as to how this is for DX signals. This will undoubtedly depend on whether or not propagation takes place by daylight. The Prolab program uses this division into ordinary and extraordinary wave and should be able to state this for DX.[4]

*Phase Selectivity in DSP
Equals Directivity*

What can DSP do with a frequency component based on the phase difference? The most obvious thing is to pass 100% of those frequency components with a phase difference that corresponds with the desired direction. All other components are not passed, but suppressed 100%. You thus create a very effective directivity. In practice, you must state two boundaries between which the phase difference of the frequency component must be in order to be passed. The width of this phase window depends among other things on the constancy of the direction from which the signal arrives. The smaller this window, the more undesired signals will be suppressed and the stronger the directivity is. Phase selectivity equals directivity.

If we have one or several interfering sources, we can also opt to suppress associated phase differences. In fact, we can therefore suppress more than one interfering source 100%. For each interfering source, we must select a matching window. The smaller the window, the smaller the chance that desired signals will also be suppressed.

Everything coming outside or within the scope of a window can be suppressed 100%. We have an infinite front-to-back ratio, which can sometimes be too good. We may not hear the signals at all that we want to or should hear. We can make the level of suppression adjustable for this purpose, depending on what is desired.

We can freely select the form of the window, as shown in Figure 8. It is not necessary to pass or suppress 100% of the frequency components that have a phase which comes within a window. We can make this dependent on the phase difference with respect to the middle of the window. The window does not have to be continuous. We may also place a notch in the window against interference that shows about the same phase difference as the desired signal.

The windows are controlled manually, just like a rotator. The center of the window is set to the corresponding desired direction. Depending on whether the source in that direction is the desired signal or an interfering signal, the signal is either suppressed or passed. The width of the window is adjusted for best readability.

Overlap in Frequency

In the beginning, it was assumed that the frequency components of various sources do not overlap one another. In fact, there is, of course, always a chance that one or more components do overlap. This possibility is not constant but depends on the coincidental frequency content of the sources and the frequency resolution in the digital processing. The consequences depend on the amplitude ratios and the width of the phase window. The source delivering the strongest contribution at that moment has the strongest influence on the net phase difference of the total of both overlapping components.

If the contribution from the desired source is strongest, the component will be considered desirable. The contribution from the interfering source will then be unjustly passed, based on the phase difference. This will manifest itself in a lower suppression of the interfering source.

If the contribution of the interfering source is strongest, the component will be considered interference. The contribution of the desired source will then be unjustly suppressed based on the phase difference. If we make the window very small, we can hear this with speech in combination with noise as a loss in the high tones. The high tones are normally lower in amplitude. Weaker components of a desired source are generally less relevant. With the combination of several overlapping frequency components and whereby the contribution of the interfering source to these components is stronger than those of the desired source, we will miss much of the desired source. Depending on the coincidental overlap, fragments occur whereby the desired signal fails partially or completely.

Noise Sources

The assumption is that every frequency component comes from one single source. This assumption is always disrupted by the presence of noise, as noise exists over the entire bandwidth.
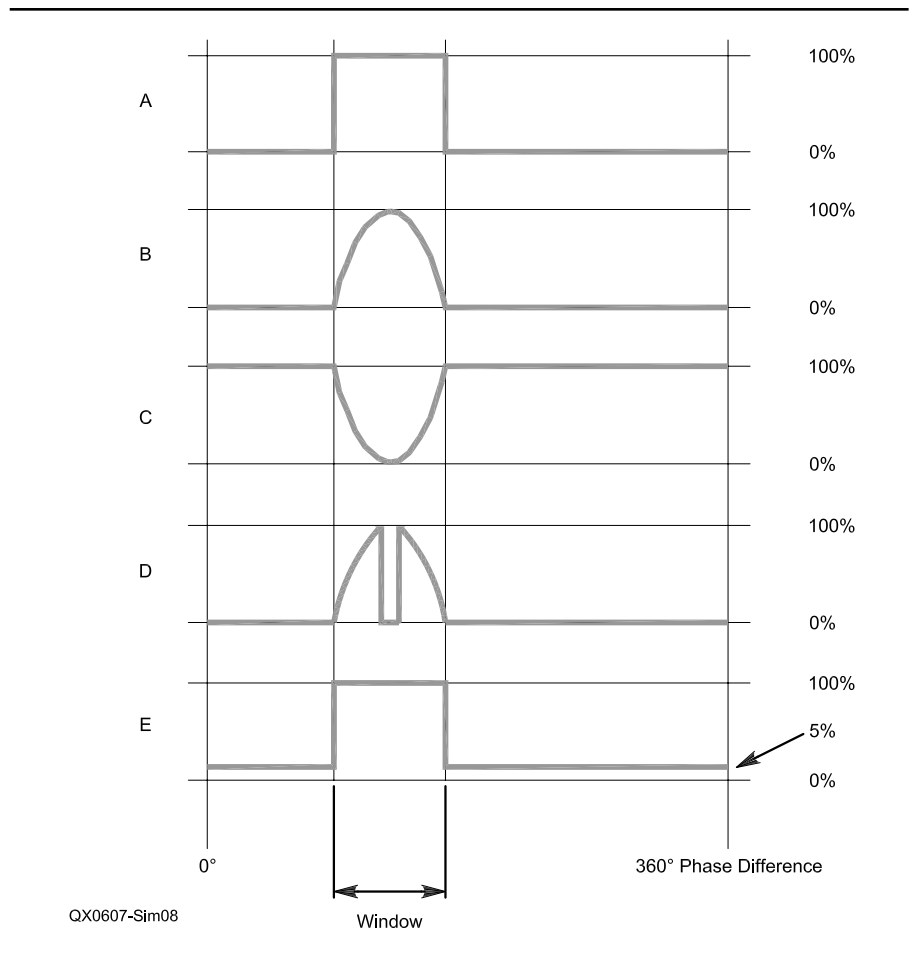

Figure 8 — In the window A) pass 100% B) pass depending on phase difference C) suppress depending on phase difference D) pass with notch e) pass 100% with 95% suppression outside the window.

We can distinguish four noise sources, man-made noise, atmospheric noise, antenna preamplifier noise and receiver noise. Man-made noise can be viewed by the large number of sources as random noise unless it concerns a distinct dominant local source.

The noise from both antenna preamplifiers and the receiver noise from both receivers are uncorrelated and can therefore have any random phase difference between 0° and 360°. Atmospheric noise and man-made noise are difficult to distinguish. Both antennas receive this noise with a difference in the time of arrival, and corresponding phase difference. The maximum possible phase difference depends on the distance between the antennas and the wavelength, as is the case with a normal desired signal. Each phase difference is also associated with an arc of directions. We therefore do not see this noise over the entire 360° but with a certain distribution over a part of this, as shown in Figure 5.

With the antennas at right angles to one another, man-made noise (ground wave) will be around 0° and 180°. Atmospheric noise can be circularly polarized and have a dominant phase difference around −90°. Linearly polarized atmospheric noise can also be distributed over the entire 360° (variable linear polarization), however.

Consequences of Noise

Were there not any noise, we would be able to determine the phase difference very accurately for each frequency component. In practice, however, each frequency component consists of a contribution of the desired or undesired signal and a contribution of the noise. This noise causes phase noise. The maximum deviation in the phase difference resulting from the noise increases as the signal gets smaller with respect to the noise. Therefore, suppression of a weaker interfering signal demands a wider phase window. If an interfering signal is roughly as strong as the noise, the need to suppress it will also be lower. Parallel to this is the fact that a stronger interfering signal needs a smaller window to be suppressed. So this works out the right way.

As the signal-to-noise ratio decreases, the achievable directivity decreases as a result of the necessary wider phase window. This is a disadvantage for weaker desired signals, as you would require of all things, stronger directivity for this.

For local (normally weaker) interference, it is possible that one of the antennas picks up the interference at a lower level than the noise level. One receiver then receives frequency components from this interference and the other only noise components. The phase difference is then dominated by these noise components. The result is that the interference components are spread out over 360°. The result is that interference with regard to phase difference is similar to noise.

Noise Reduction

Antenna amplifier noise and receiver noise can be found over the entire 360° phase difference. Atmospheric noise and man-made noise appear over a smaller range, defined by the distance between the antennas and the operating frequency. The amount of noise we pass is determined by the part we pass via the phase window. The width and form of this window determines how much noise is effectively passed. But also the amplitude ratios of the various noise sources is of influence, as it determines the distribution of the phase difference over 360°.

For the antenna amplifier noise and the receiver noise, reduction can increase considerably. A phase window of 45° will already show a reduction of 9 dB. A local interfering source, which is received by only one of the loop antennas, will also undergo similar reduction. Atmospheric and man-made noise will be reduced less. This noise is not distributed over the entire 360° but is concentrated within a smaller area.

Hardware

Figure 9 shows the total system. Both antennas are connected to the FT1000D. The FT1000D has two synchronized receivers, a main receiver and a sub receiver. Unfortu-
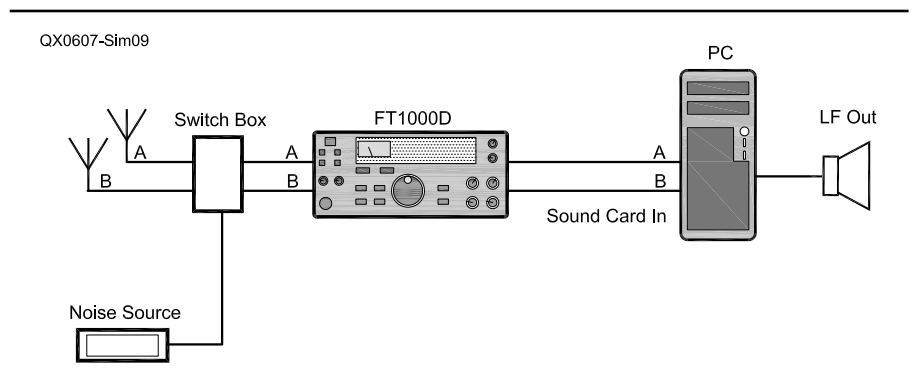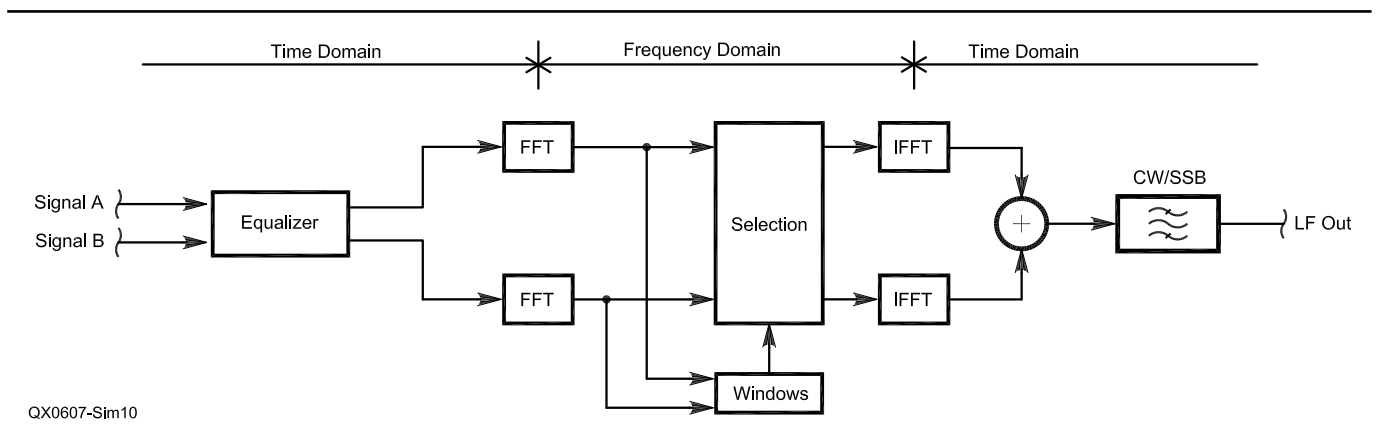
Figure 9 — The total system.

Figure 10 — Block diagram of the digital signal processing.

nately, both receivers are not identical. In the DSP we have to take into account the fact that the frequency and phase characteristics are different and that the phase difference between both receivers is constant in time, but unknown. A modification has been fitted in order to tune in both receivers via the main tuning button.[5]

The audio signal from both receivers is acquired via the sound card. No extreme specifications are demanded of the sound card as the receivers already carry out the AGC. A disadvantage is that software buffering is required on input and output for the uninterrupted processing of sound. This leads to an extra non-functional delay of 0.2 seconds. Apart from the fact that this is not ideal for tuning, it handicaps QSK telegraphy. For real-time processing, a 3-GHz Pentium PC is required.

The noise source is used to equalize both receivers in the DSP regarding frequency characteristics. For this, we connect the noise signal to both receivers at the same time.

An adaptive filter then corrects the audio signal from the sub receiver.

DSP Functionality

The key to the digital signal processing lies in the FFT (fast Fourier transform).[6] See Figure 10. The FFT can be seen as a large number of bandpass filters. With this, we can divide the audio band from 0 Hz to 4 kHz in 256 bands of 15.625 Hz. Because the bandpass filters overlap each other somewhat (spectral leakage), the net frequency resolution becomes about 30 Hz. If we increase the FFT frequency resolution, this will be associated with a longer time window. The frequency content of the sources varies in time and the chance of a contribution from an interfering source into one of these bands will remain roughly the same. But the average amplitude of this contribution decreases accordingly. So, a higher frequency resolution gives progressively better results. A four times higher resolution (7.5 Hz) is feasible at the cost of an increased time delay in the processing.

Apart from amplitude, the FFT also calculates the phase of each frequency component and it is therefore simple to calculate the phase difference. In the "windows" block, the software determines if the phase difference is within a window, and is calculated, depending on the form and function of a window, how strong the respective component must be passed or suppressed. The "select" block will calculate the final strength of the frequency components. The "IFFT" (inverse FFT) block converts it into a signal in the time domain.

With the "equalize" block, at the beginning of the block diagram, we can correct for differences in both receivers using an adaptive

filter. With the noise source connected to both receivers at the same time, an adaptive filter adjusts the frequency characteristic in such a way that the sub receiver signal equals the signal from the main receiver.

The CW-filter has a somewhat hidden function. We decide per unit of time whether a frequency component may be passed. A passed frequency component during this unit of time has a constant amplitude output and is hard switched on and off. This produces the same effect (key-clicks) as that of a hard switched CW signal. With this CW filter, we reduce these "clicks." With the bandwidth used for SSB, this phenomenon is not audible.

Programming in Matlab/Simulink

Matlab is a most suitable environment in which to develop a DSP application.[7] It demands the necessary knowledge in the areas of DSP and mathematics, so it is not really meant for beginners. The block diagram in Figure 10 was programmed with Simulink. Simulink uses ready-made blocks in which basic functions are pre-programmed. This is handy as it is not necessary to reinvent the wheel. Another significant advantage is that Matlab is capable of producing an executable program, which increases processing speed considerably. This translates into the possibility of carrying out more complex processing in real time. It would be a little too much to explain the processing in greater detail in this article, but I will quote some figures. There are 4 FFTs and 4 IFFTs, each 512 (up to 1024 or 2048) samples long executed with a sample frequency of 8 kHz. The adaptive filter is of the least-mean-square (LMS) type.

Figure 11 shows a screenshot. The control panel is shown in the bottom half and is programmed using Labview.[8] For two windows, the phase difference (−180° / −90°) and the window width (0° / 30°) can be adjusted. There are switches to equalize both receiv-

ers, for selecting between CW or SSB, for selecting the main-receiver or DSP output, for passing or suppressing signals in the windows and for placing a notch in the window or not. There are settings for the depth of the suppression of the undesired frequency components and for an amplitude level below which frequency components (for example noise) will not be passed. Also, noise cancelling and an AGC function are implemented. An S-meter indicates the audio-signal strength and the AGC gain.

Two graphs show the phase difference of the frequency components (0 to 4 kHz) and a cumulative histogram of phase differences (upper right). There is also a setting for an amplitude level below which frequency components are not used for both graphs. The phase difference fitting the peak in the histogram (−87°) is shown separately in numerical terms. This makes it easier to adjust the windows.

Limitations

Practice shows there is only a small chance that frequency components of various sources overlap each other. The digital signal processing is based on this assumption.

This assumption also constitutes the major limitation, however. With the combination of many overlapping frequency components, whereby the contribution of the interfering source in these components is stronger than that of the desired source, the results described below are not feasible. In these cases, we can always fall back to the nulling properties of the loop antennas and to analog or digital (DSP) noise cancelling. (See Note 1.) In the DSP, we can also increase the frequency resolution from 30 Hz to, for example, 7.5 Hz. This gives progressively better results, especially for CW. The extra delay in the processing, however, is very noticeable. A 30 Hz net frequency resolution seems a good compromise.
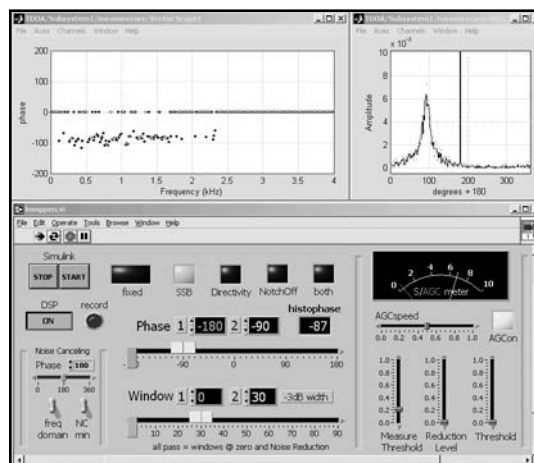

Figure 11 — Screenshot of the control panel and graphs. The control panel is shown in the bottom half and is programmed using Labview.

Despite the fact that the digital signal processing has not been carried out on the intermediate frequency level, it turns out to be very effective. The need for processing on the intermediate frequency level will only become noticeable if the interfering signal is much stronger than the desired signal. The desired signal will then be modulated too intensely by the AGC. It is better to control the AGC manually. This last limitation is not fundamental, however.

Results

The directivity achieved with DSP and the two loop antennas in the same direction is very convincing. Weak signals will be easier to copy with the noise reduction, although the useable directivity decreases with decreasing signal-to-noise ratio. Local and non-local interfering signals are easy to suppress. Even at 80 meters, and with an antenna distance of just 7 meters, stations coming from different directions can be distinguished. At 40 meters for example, it is often quite possible to separate G (west) from DL (east) stations here in PA. This gets even better on the higher bands because the distance between the antennas increases with respect to the wavelength.

For the 80 meter and 160 meter bands, the use of circularly (elliptically) polarized signals is very effective. The loop antennas are placed at right angles to one another for this purpose. If there is stable circular polarization, we can select the desired signals via the associated phase difference (–90°) resulting in a very quiet band. This is normally the case by day at 80 meters. Around and after sunset, the polarization sense on the 80 meter band may sometimes turn around and we see a phase difference of around +90°. In some cases, local PA stations distinguish themselves from DX stations at 80 meters in the sense of this circular polarization. The associated phase differences are then opposite and we can separate these stations very well. If polarization becomes linear at 80 meters after sunset, the phase difference will switch between 0° and 180°. Only local interfering sources coincidentally showing another phase difference can then be suppressed. For linear polarization, it is better to place the antennas in the same direction and to use directivity.

To what extent placing the loop antennas above one another can be used for selection on the elevation angle has not yet been tested. There should be added value from the fact that DSP adjustments are not needed as often.

Audio samples tell more than words and figures. That's why there are some audio samples on my Web site (see Note 1) that give an idea of the effectiveness of this digital signal processing.

Comparing Noise Cancelling Techniques

In principle with DSP, the same noise cancelling can be performed as, for example, the MFJ-1025 or ANC-4 perform on high frequency signals. DSP can cancel interference much better and faster using adaptive filters. When cancelling a local interfering source, two different antennas can be used.

Although a lot is possible with adaptive filters, there are limitations in the level of noise reduction achievable with two antennas used as a receiving array. In such an array, the signals from the antennas are summed after applying a complex weight to the signals. When using two antennas there is a maximum of just one interfering source or one arc of direction to be suppressed. Figure 12 shows an example of a possible 3-dimensional radiation pattern of an array of two loops, as shown in Figure 3A. The arc of nulls can be very sharp, but the beam is always broad and the achievable directivity is limited. When the distance between the loops gets smaller compared to the wavelength ($<1/8\,\lambda$), the gain decreases significantly.[9, 10, 11, 12] The sharp null makes it difficult to suppress interfering sky wave signals, because on shortwave the direction of arrival is constantly changing over time. The advantage is that such an array can operate if two or more signals share the same frequency.

The frequency domain processing using time difference of arrival or the phase difference for selecting frequency components as described in this article, behaves much differently. Beams and nulls are in arcs of direction as shown in Figure 4, but very sharp beams and very broad nulls are possible. Also, multiple beams and multiple nulls are possible. The achievable directivity is only limited by the signal-to-noise ratio. The gain is not affected when the distance between the loops gets smaller compared to the wavelength. A further advantage is that it is possible to profit by elliptical polarization.

All in all, each technique has its own advantages and disadvantages. There is not one single best technique. The best you can do is have all techniques at your disposal. Digital Signal Processing makes this feasible.

Epilogue

A lot more is certainly possible in the field of DSP. A challenge is carrying out this
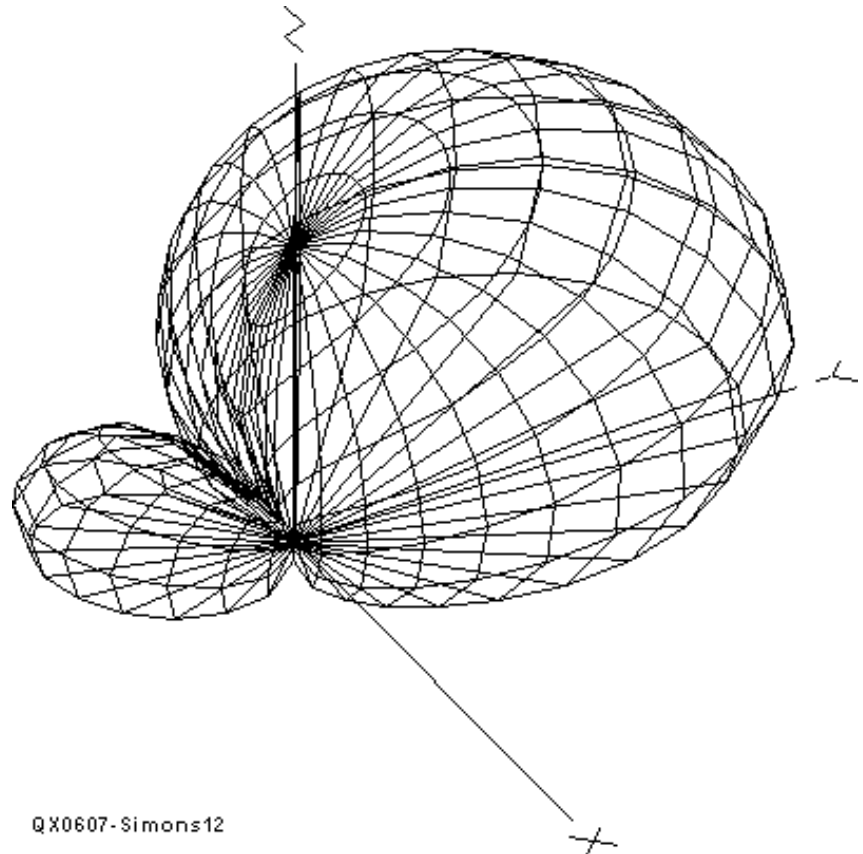

QX0607-Simons12

Figure 12 — A three-dimensional radiation pattern of an array of two vertical small loops, using summing of signals.

processing on the intermediate frequency level of two synchronized receivers. By "synchronized," I do not just mean exactly the same frequency, but also a fixed phase difference (phase coherent). With the FT1000D, the phase difference is changed after each frequency tuning and also after each receive/transmit/receive cycle. The Flexradio concept is maybe the most promising concept.[13]

Another challenge is extending the number of antennas and receivers. This would mean we can make a better selection of the desired direction.

Instead of loop antennas, other antennas, such as ground-planes or dipoles can be used. The associated radiation patterns each have their advantages and disadvantages. When making this choice, remember also the necessity for good balance and the level of mutual coupling between the antennas.

For reducing BPL (Broadband over Power Line) interference, this set-up is certainly no solution. BPL fills the entire band and, in case of mass use, comes from every direction.

Notes
[1] PA0SIM (audio samples) on the Internet: **http://home.plex.nl/~jmsi/**.
[2] R. Dean Straw, N6BV, *The ARRL Antenna Book*, 20th edition, The American Radio Relay League, 2003, p 3-11. *The ARRL Antenna Book* is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 9043. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.
[3] Dourbes: **http://digisonde.oma.be/**.
[4] Prolab: **www.spacew.com/www/proplab.html**.
[5] Modification FT1000D: **www.angelfire.com/md/k3ky/page61.html**.
[6] See the discussion about The Fourier Transform at: **http://users.rowan.edu/~polikar/WAVELETS/WTpart2.html**.
[7] Matlab/Simulink: **www.mathworks.com/**.
[8] Labview: **www.ni.com/labview/**.
[9] D. Smith, KF6DX, "Introduction to Adaptive Beamforming," *QEX,* Nov/Dec 2000, pp 50 – 55.
[10] B.Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
[11] J. Devoldere, ON4UN, *Low-Band DXing*, Fourth Edition, The American Radio Relay League, 2005, Chapter 7. *Low-Band DXing* is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 9140. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.
[12] D. Smith, KF6DX, *Digital Signal Processing Technology*, The American Radio Relay League, 2001, Chapter 12. *Digital Signal Processing Technology* is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 8195. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.
[13] Flexradio: **www.flex-radio.com/**.
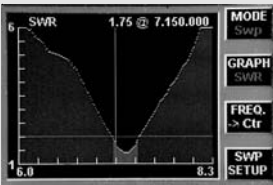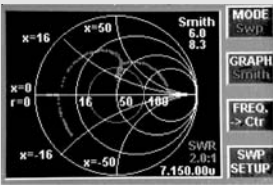
Jan Simons, PA0SIM has been a licensed radio amateur since 1975. He works as an analog and digital electronic engineer for a printer and copier manufacturer in the Netherlands. Designing mixed signal ASICs is a part of his job. As a radio amateur he especially enjoys working DX and CW on shortwave. Dealing with QRM is a real challenge when living in a residential location. Improving reception on shortwave has been a major activity for the last few years. QEX–

Practical RF Soil Testing

Have you wondered what ground-conductivity value to use when you model an antenna design? This article describes a measurement technique that can provide you with an answer specific to your location.

Eric von Valtier, K8LV

Antenna modeling has become a new area of Amateur Radio experimentation as well as a highly useful tool for designing antenna systems. Modern hams prefer to try their new antennas out on the computer before actually building them. This approach can save many hours of wasted effort and material, primarily by showing what designs are *not* going to work well. LF antennas are particularly troublesome because of the influence of the ground on wave generation and propagation, and past efforts have all used a variety of tricks and magic rules to account for the ground.

Every user of antenna modeling software soon realizes that the parameters of the ground itself are highly influential on antennas, but little is really known about how to accurately specify the ground material. There are "standard" numbers in use, some from reputable sources and others from questionable sources. I am certain that at one time or another we would all like to be able to measure our own ground characteristics, and this article is intended to satisfy that need. It will show how relatively accurate measurements can be made with a simple setup accessible to many hams.

Some Basic Physics

Electrical conductivity is one of the most basic of all electrical quantities whose effects are readily manifested in the macroscopic world. Virtually every radio amateur, regardless of their degree of electronics knowledge, has some familiarity with "ohms" of resistance in some context. In this article that context will be centered

521 Leicester
Plymouth, MI 48170
evonvaltie@aol.com
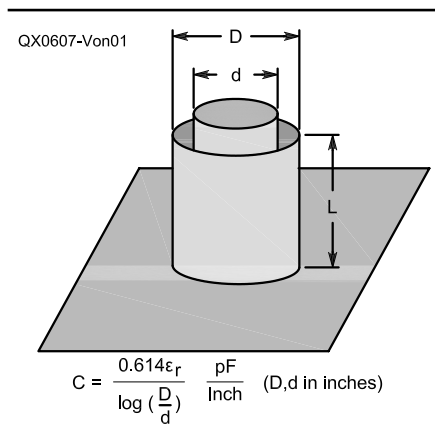
Figure 1 — This diagram shows a test cell that is based upon coaxial capacitor geometry.

$$C = \frac{0.614\varepsilon_r}{\log\left(\frac{D}{d}\right)} \quad \frac{pF}{Inch} \quad (D, d \text{ in inches})$$

on antennas and their environment. Unfortunately, there are some subtle complexities appearing in the theory of soil conductivity. In particular, most of all radio and electronics practice involves electrical conductivity occurring as a result of the motion of electrons in metallic substances, called *conductors*. The theory of metallic conductivity is well understood and documented and presents little or no problems in analysis of circuits, wires, components, and so on. The most significant complication that arises in metallic conductors is skin effect, which causes RF currents to avoid flowing within the interior of conductors and concentrate toward the surface. The resulting resistance to current increases, but in a way that can be readily handled as a simple "correction" to the dc resistance.

Conductivity arises in any substance that contains electrons that can be freed from their bound state in the base medium, and

allowed to move freely under the application of an electric field, producing electron flow: electric current. Many liquid solutions containing dissolved solid substances are prime examples of this "liquid conductivity," where the current is predominately due to ionized solute molecules. It is generally much lower than solid conductivity, but still of considerable significance in certain cases. One important case is the conductivity that occurs in the soil composing the Earth ground that we live on and build antennas upon, and its effect on radio waves.

The theory of liquid conductivity, also called *ionic conductivity,* is not nearly as well understood as solid-state conductivity. In particular, RF conductivity in liquids is a highly complex issue that will not be dealt with much here. The position taken will be the orthodox one, that skin effect and RF resistance occur in liquid conductors, and that they can be properly accounted for by a *complex permittivity* that can be accurately measured empirically. The resulting values are then applicable in Maxwell's equations as usual to perform field theory calculations.

In practice, the complex permittivity, the basic quantity that relates electric current (density) to the electric field in any medium, is expressed in terms of its two components σ and ε, which are defined by the formula:

$$\mathbf{J} = (\sigma + j\omega\varepsilon)\,\mathbf{E} \qquad \text{(Eq 1)}$$

The real component, σ, describes the conventional current due to electron motion, and the imaginary component, ε, describes the displacement current. They are related to the more commonly used parameters of resistance and dielectric constant by simple inverse relationships (reciprocals) but for EM field work the defining format of Equation 1 is generally preferred. Users of antenna modeling programs are familiar with their presence as input data required

for antenna calculations, especially those involving antenna structures in proximity to the Earth.

Antenna Modeling

The permittivity values can be very influential in the results obtained from an antenna calculation. This should come as no surprise since it is deeply embedded in the "fourth" Maxwell equation:

$$\nabla \times \boldsymbol{H} = (\sigma + j\omega\varepsilon)\,\boldsymbol{E} \qquad \text{(Eq 2)}$$

where the "current" term has been written in terms of the complex permittivity, which contains a real and an imaginary term, each with the dimensions of siemens (reciprocal ohms). The imaginary term is expressed frequently as a value relative to its value in free space:

$$\varepsilon = \varepsilon_r\,\varepsilon_0$$
$$\varepsilon_0 = 8.87 \text{ pF/m.}$$

The relative permittivity is also known as the *dielectric constant* for the material in question. Equation 2 is viewed by most EM practitioners as the "source" term for production of waves, and it is obvious from its form that the values of the permittivity components will play a key role in wave motion. All wave motion is explainable by solving this equation simultaneously with the remaining three Maxwell equations. This, of course, is the basis for all antenna calculations and programs that perform them.

There are numerous rearrangements of Equation 2 that result in the appearance of modified versions of σ and ε but the added complexity is not necessary, since the antenna modeling software typically requires direct values of these quantities as primary input. Hence, the focus here will be on measuring them as quickly and directly as possible. There is a small amount of unit juggling in the method described, but it has been reduced the to the minimum possible.

Measurements of Permittivity

The measurement of permittivity can take several different paths, and I have tried several of them in an effort to find a practical method for simple lab setups. I will discuss only one of these methods in this article, as it has proven to be simple enough to be used by anyone who has to access to an impedance meter. It is based upon the electric field relationships within a confined space, filled with the soil to be tested. In particular, the interior region of a coaxial capacitor is well suited for this, since it confines the field to a known, and calculable, region with minimal error due to fringing at the open end.

The generic coaxial capacitor shown in

Figure 2 — Here is a photo of the test cell packed with a sample of soil under test, set up for a frequency sweep on the impedance meter.

Figure 1 can be analyzed using electrostatics only, as it is assumed that the length of the capacitor is small enough to produce negligible longitudinal phase shift. The following shows how measurements of the full permittivity can be obtained in terms of the current into the capacitor for a given applied RF sine wave voltage from a generator. The current is related to the voltage by the formula:

$$\boldsymbol{I} = \boldsymbol{V}\,\boldsymbol{Y} \qquad \text{(Eq 3)}$$

where all quantities in bold type are understood to be vectors (complex values like $R + jX$). This formula defines the complex admittance, which is the reciprocal of the complex impedance.[1] The admittance, $\boldsymbol{Y}$, is related to the capacitor geometry as follows:

$$\boldsymbol{B} = \omega\,\varepsilon_r\,C \qquad \text{(Eq 4)}$$
$$\boldsymbol{G} = \sigma\,C\,/\,\varepsilon_0 \qquad \text{(Eq 5)}$$

The value of C is found from the well-known formula for a coaxial capacitor with outer and inner electrode diameters of D and d:

$$C = 0.614\,\varepsilon_r\,/\,\log\,(D\,/\,d) \text{ pF/inch} \qquad \text{(Eq 6)}$$

When the capacitor is connected to a complex-impedance meter, it will read a value for $\boldsymbol{Z}$, which can then be inverted to produce a value of the complex admittance $B + jG$. Applying this in Equations 4 and 5, which are then rearranged to produce final values of $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$, gives the actual computational formulas:

$$\boldsymbol{\sigma} = \varepsilon_0\,\boldsymbol{G}\,/\,C \qquad \text{(Eq 7)}$$
$$\boldsymbol{\varepsilon}_r = \boldsymbol{B}\,/\,\omega\,C \qquad \text{(Eq 8)}$$

Thus, an accurate measurement of these values can be made at any frequency. For reference, the conversion from impedance to admittance will be repeated here in both rectangular and polar form. If the impedance meter displays the polar vector impedance in the format $Z \angle \theta$:

$$G = (1\,/\,Z) \times \cos\,\theta \qquad \text{(Eq 9A)}$$

[1]Notes appear on page 49.

$$B = -\,(1\,/\,Z) \times \sin\,\theta \qquad \text{(Eq 9B)}$$

For a meter displaying rectangular components $R + jX$:

$$\boldsymbol{G} = R\,/\,(R^2 + X^2) \qquad \text{(Eq 10A)}$$
$$\boldsymbol{B} = -X\,/\,(R^2 + X^2) \qquad \text{(Eq 10B)}$$

A Soil Test Cell

A practical setup for applying the above can be built from common materials with little effort. I will describe one such effort in some detail, and then present some actual results.

Figure 2 shows one test cell in operation at K8LV. The outer electrode is a short length of 3-inch aluminum duct and the inner electrode is a length of 2-inch PVC pipe. The latter is wrapped with aluminum duct tape to build its OD up to a diameter that leaves about a half inch spacing between inner and outer electrodes. The two are secured to a base of circuit board material via RTV cement, and the outer electrode is grounded to this base. The conductive wrap on the inner conductor is truncated about a quarter inch above the base to minimize internal fringing.

The cell is packed with the soil under test (SUT) up to the top of the outer conductor. With short leads connected to both electrodes, the cell is connected to the impedance meter and readings taken at all desired frequencies. These $\boldsymbol{Z}$ readings are then converted to permittivity values using Equations 9A and 9B and then Equations 4 and 5.

The final test cell can be calibrated if desired. In my experience the calculated value of the cell was always very close to the calibrated value, and rendered calibration unnecessary. The following calibration procedure can be used if desired, however.

The cell is filled with distilled water and then isopropyl alcohol, both of which have known values for ε. Their dielectric constants are:

$\varepsilon_r = 80$ (water)
$\varepsilon_r = 20$ (alcohol)

The measured impedance values for these samples should be pure negative reactances, from which the "air" value of the coaxial capacitor can be directly calculated as:

$$C = 1\,/\,(\omega\,\varepsilon_r\,X_c) \qquad \text{(Eq 11)}$$

The only problem with this procedure is that the cell must be made liquid-tight to prevent leakage, which is not really necessary for soil samples.

The dimensions of the test cell are not critical. It is desirable to keep the impedance readings in the range of tens to hundreds of ohms, where most meters have maximum accuracy. The size shown here is a good compromise between field containment (fringing fields outside of the capacitor) and inductance, which would result from

excessive lengthening. It is desirable to use a wide spacing between electrodes to minimize the effects of small holes or lumps in the SUT, but not so wide as to result in low values of C.

The test cell shown in Figure 2 has a C_{air} value of 18.2 pF, becomming several hundred pF when packed with soil samples.

When packing the cell I first screen the sample by hand to remove small stones and pieces of foreign matter, but it is important to not handle it excessively. It should be processed as soon as possible after extraction to avoid water loss due to evaporation. It should be packed into the cell in a manner which best reproduces its consistency in nature.

Some Results

I have made numerous measurements with this setup over the last several years. In general, the results have been quite gratifying. These measurements are generally repeatable for samples from the same area. They agree well with most published data for σ but the data for ε raises questions, which will be discussed shortly.

Figure 3 shows a typical frequency sweep of a sample from K8LV. The soil is brownish sandy midwestern loam, with moderate moisture content. Samples have been taken at depths of 1 to 8 feet and they show remarkably small variation. It is best to take samples from at least a foot or two depth, where the moisture level has begun to stabilize. Bear in mind that the skin depth in average soil is approximately one-third of the square root of the RF wavelength, which is about 4 meters on 160 m.[2]

Samples from stations in our local area have all produced measured permittivity values in the range shown in Figure 3, which is encouraging, and lends confidence to the test setup and procedure. After accumulating a number of such data, I became very interested in comparing it with other regions.

Figure 4 is data taken from a well-known east coast station, W1MK. I contacted Rob Lahlum because he has a long record of exceptional performance on 80 m while using antennas that are somewhat modest by "big station" standards. His samples were sent to my lab via air freight in sealed containers and the tests were performed within 24 hours of their extraction from the ground. Several samples were taken at different locations on his property, and the results were all amazingly similar. They show, as was expected, a higher than normal conductivity. "Normal" here is synonymous with "Midwest," my obviously biased reference point. When this data is used in comparative calculations on typical vertical antennas using the "home" values, they
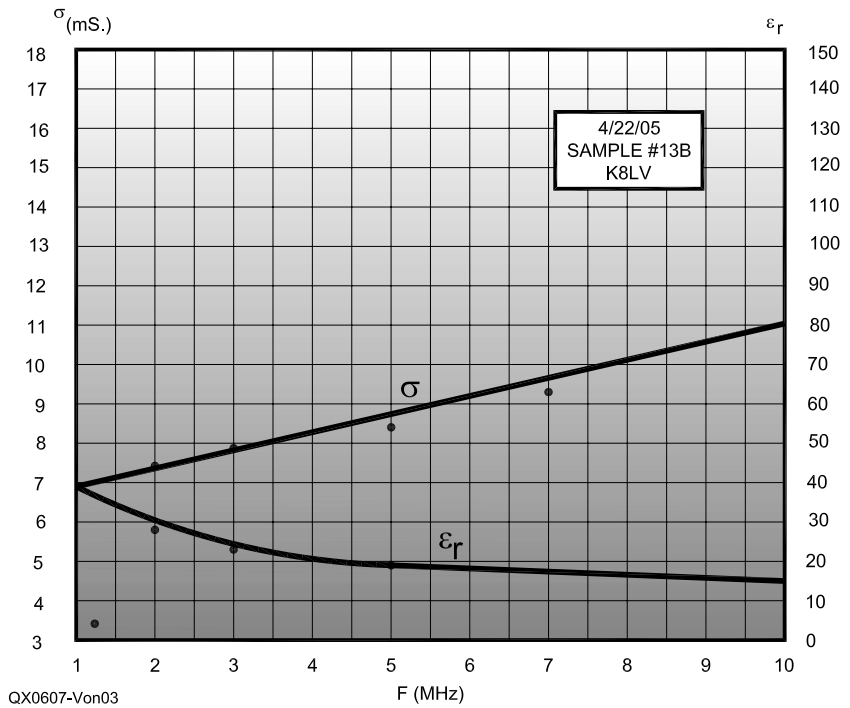
Figure 3 — This graph shows the permittivity of typical Midwestern USA soil. Solid lines are best-fit curves to the data points.

consistently shows increased gains of 1 to 2 dB, with corresponding increases in radiation efficiency.

Also notable in this data is the fact that it falls quite short of the 30 mS. value cited as "very good" soil in some of the NEC programs. In fact this soil may be almost as good as it gets, considering that it comes from a location close to the ocean on the east coast. It will be illuminating to see results from various parts of the country (and the world) by others who make measurements similar to those described here.

Some interesting additional tests were performed to verify the notion, stated previously, that this conductivity arises from the liquid conductivity. Several samples of soil were frozen and measurements taken. The resulting values of σ dropped to a small fraction of a milliSiemen and ε dropped by factors of 10 or more. When the samples were thawed the readings always returned reversibly to their ambient values. The frozen ε values always fall within the range of pure silicon dioxide, the major constituent of sand, a major ingredient of most soils.

The frozen soil data reflects the permittivity of the solid material in the soil,

but the relevance of that data is not great. It simply supports the assumption of the ionic conductivity that results in the measured values of σ.

Alternate Measurement Methods

I have not attempted to correlate these measurements with dc measurements. I believe the latter to be somewhat irrelevant and of little value. Their primary problem is that even if we assume that an accurate dc measurement can be made using multi-probe techniques, the relationship between RF and dc resistance is not well enough known for a substance of such physical complexity as soil to properly correct these readings. In addition, that method provides no data on epsilon, which is shown in my data to vary considerably from known "assumed" values. When the sample is placed in a carefully controlled and known environment, however, the voltage and current can be accurately measured to produce permittivity data of estimable accuracy. This *can* be done easily, as has been shown here.

An alternate approach has been discussed by some, including me, which involves measuring the resonant frequency shift of a dipole radiator located at ground level. The appeal of this method is that it may tend to
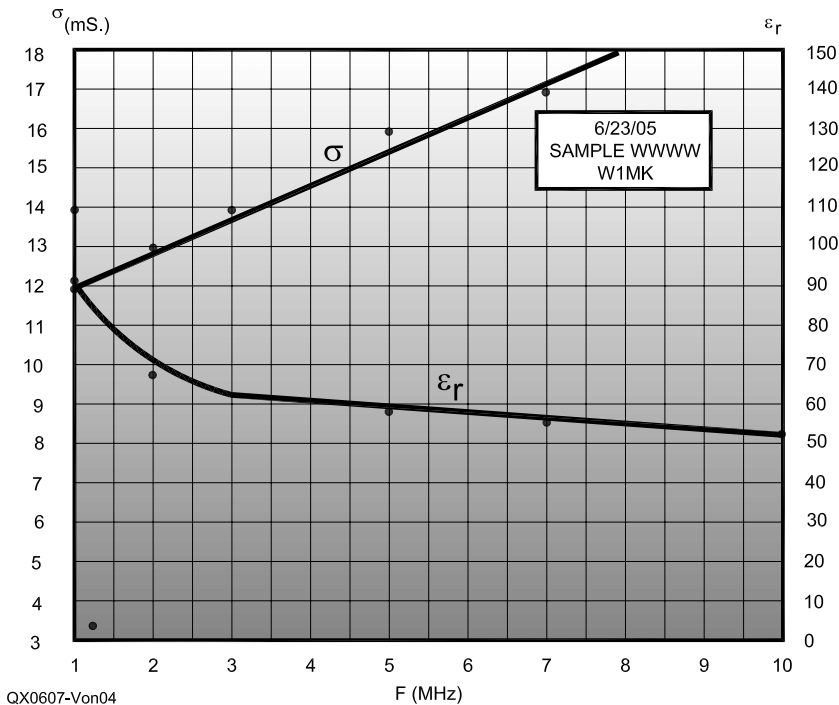
Figure 4 — This graph shows the permittivity for high-conductivity East-Coast soil from a sample taken by W1MK.

better account for the non-homogeneity of the ground soil and water content. It will tend to average all of these out, but it is certainly not clear how to properly use the data. At the present, there is no useful theory of antennas over inhomogeneous ground. Computational regimes such as NEC are just able to handle one discontinuity in the antenna space — that between free-space and the Earth. It is theoretically possible to extend the Sommerfeld ground theory, the core of the NEC method, to multiple layers, but with enormous complexity. When and if such code will appear is not clear at this time. For the present, the two-layer regime of NEC is providing very useful results for ground based antennas, which were simply not available more than ten years ago. We must expect a certain degree of inaccuracy due to the inherent difficulties in characterizing the ground medium. The contents of this article are meant to help minimize that error as much as possible.

Notes

[1]Like many calculations in RF applications involving wave motion, admittance parameters are frequently more convenient to work with than their matching Z parameters.

[2]The mythical "surface layer" does not exist in fact, and the RF fields of verticals penetrate 10 to 20 feet into the ground on 160 and 80 m.

Eric von Valtier, K8LV, is an engineer and physicist with advanced degrees in electrical engineering, nuclear engineering, and physics from the University of Michigan. He was first licensed in 1951 as WN8IRO, spent his entire teen years obsessed with ham radio, and went QRT from 1960 to 1998, returning to ham radio as a "retirement" activity. His professional career was spent as an independent design and R&D engineer, during which he produced over 100 product designs that were successfully marketed, with several patents issued. He is equally experienced in both analog and digital electronics and has written over a half-million lines of code covering numerous applications from commercial database access to scientific computation. He enjoys building equipment for his station using new and experimental concepts, and pursuing LF DX using antennas that perform beyond the expectations of a QTH which cannot "be taken very seriously for DX purposes!"

QEX-

Crystals

Analyze the tuned circuit that is a crystal, and learn about ways to use these small bits of quartz to control an oscillator frequency.

Andrzej Przedpelski, KCØCWK

We use crystals in many everyday applications: radio equipment, watches, PCs and other circuits as filters and frequency determining devices. Thus, it may be interesting to examine their properties, especially since some misconceptions may exist. A crystal is basically a tuned circuit. It can be either series or parallel resonant. To illustrate, a typical 2 MHz crystal has an approximate equivalent circuit as shown in Figure 1. The crystal has L, C and R characteristics, as shown. Stray capacitance, including the holder capacitance, is represented as C_o.

7260 Terrace Place
Boulder, CO 80303
kc0cwk@arrl.net

The transfer function, g(f), from terminal A to terminal B is shown in Figure 2. While it may look somewhat intimidating, it is mainly the frequency response of a series tuned circuit in parallel with a capacitor.

This equation can be derived as follows. The series circuit impedance, X_S, is $X_L + X_C + R$, where X_L and X_C are the reactances of the inductor and the capacitor, respectively. This impedance is in parallel with the impedance of C_o, X_P, resulting in the overall transfer function $g(f)$, using $(X_S)(X_P) / (X_S + X_P)$.

This can be better seen by plotting this function, as shown in Figure 3. It is apparent that it consists both of a series and a parallel resonant circuit.

At series resonance, it is resistive, with a value of R. At parallel resonance, the LC series circuit is inductive (above series reso-

nance) and resonates with the stray capacitance, C_o.

Let's look at a typical application in a crystal-controlled oscillator. Basically, an oscillator consists of a frequency selective network and an amplifier to provide the necessary gain to sustain oscillation.

Series Mode Oscillator

Let's consider the case where the crystal is operating in the series mode. This is shown symbolically, in Figure 4.

To oscillate, the loop gain has to be more than unity, and the loop phase shift has to be 0°. Let's assume a perfect amplifier, with no phase shift. While Figure 3 shows why the oscillation should occur at series resonance (minimum insertion loss), it does not explain why the frequency stability

Figure 1 — Equivalent circuit of a 2 MHz crystal.

$$g(f) = \cfrac{1}{\left[\cfrac{1}{(R + j2\pi fL) - \cfrac{j}{2\pi fC}}\right] + \cfrac{j}{\left(\cfrac{1}{2\pi fC_o}\right)}}$$

Figure 2 — Transfer function, g(f)
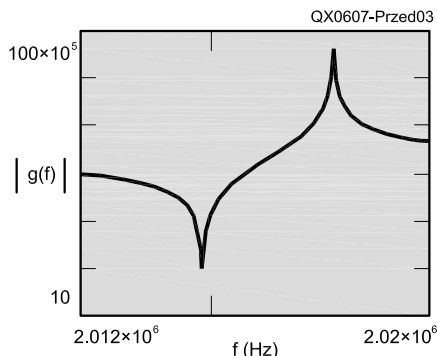
Figure 3 — Plot of the absolute value of the transfer function, g(f).
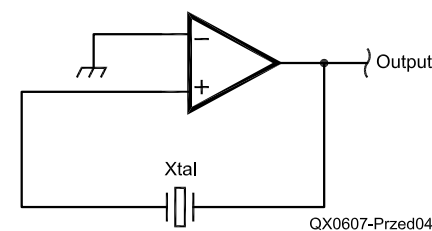
Figure 4 — Typical crystal oscillator with the crystal in series resonance.
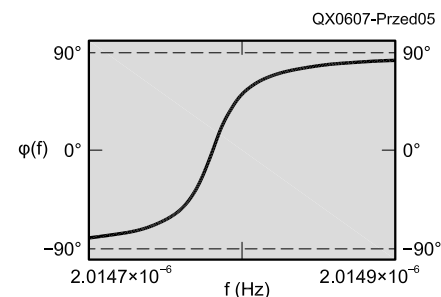
Figure 5 — Phase shift of the transfer function, g(f).

should be good. For this explanation we have to look at the phase shift through the crystal, φ, shown in Figure 5.

It is apparent that 0° phase shift occurs at 2.01478 MHz. The slope of this curve at resonance is 3.74°/Hz. Thus, even if the external circuit phase is changed by 37.4°, the frequency will change only 10 Hz to reestablish the necessary 0° overall loop phase shift. This can also be confirmed by examining the Q of the crystal at series resonance:

$$Q = \frac{1}{2 \times 2.0148 \times 10^6 \times R \times C} = 6.5827 \times 10^4 \quad \text{(Eq 1)}$$

$$\Delta\varphi = \frac{360}{\pi} \times \frac{Q}{2.0148 \times 10^6} = \frac{3.74°}{Hz} \quad \text{(Eq 2)}$$

The frequency of oscillation can also be confirmed:

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{C\,L}} \quad \text{(Eq 3)}$$

$$f = 2.0148 \times 10^6 \ \text{Hz}$$

To complete the picture, the phase shift at parallel resonance is also 0°, as is shown in Figure 6.
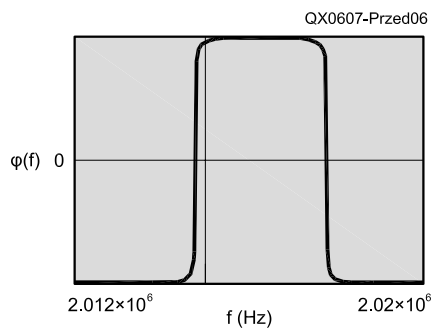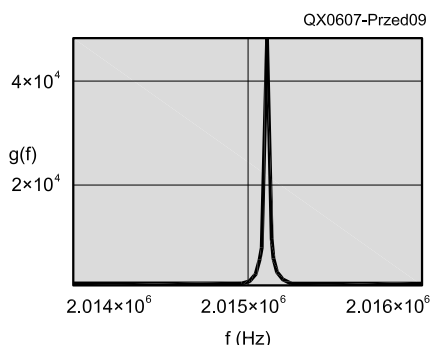
Parallel Mode Oscillator

An oscillator with the crystal in parallel resonance can be represented as shown in Figure 7. The external capacitor, C_p, is usually specified as 32 pF at this frequency.

In this configuration, the crystal (looking like an inductor) forms a parallel resonant circuit with external capacitor, C_p. Figure 8 confirms that the crystal looks like an inductor between series and parallel resonance of the crystal.

With the parallel capacitance, C_p, the parallel tuned circuit resonates at 2.015116 MHz. This is apparent from the impedance across the circuit, shown in Figure 9.

The other condition, zero phase shift, is also satisfied, as shown in Figure 10.

Overtone Versus Harmonic Operation

While fundamental crystals have been made to operate above 300 MHz, they are fragile and difficult to make. Two possible solutions exist: harmonic or overtone operation. Harmonics of the fundamental frequency, f_o, at multiples of that frequency can be used. Even with extensive filtering of the desired harmonic, there are many undesired outputs at the fundamental and the remaining harmonics. A better solution is the use of overtones. As with music, the overtones are not exact multiples of the fundamental frequency. A model showing the overtone operation is discussed by WaiTak P. Lee and is shown in Figure 11. (Only the fundamental response and the first two overtones, are shown.)

Only odd overtones are possible. To make the crystal oscillate at the desired overtone, the unwanted frequencies have to be suppressed. Usually, a simple tuned circuit is enough. The main advantage of the overtone versus harmonic operation is the reduction of spurious outputs. The fundamental and even harmonics are absent. Only the harmonics of the overtone used are present. Operation at the 9th and 11th overtone has been observed. Such high overtone circuits, however, are not very reliable.

Crystal Selection

When selecting a crystal for a specific application there are several factors to consider.

1) Phase noise. Lately, the need for low phase noise signal sources has become evident for critical applications. High Q crystals are desirable for this application. The phase noise close to the carrier is approximately proportional to $1/Q^4$ (John R. Vig). The associated active circuit should

Figure 6 — Phase shift at series and parallel resonance.
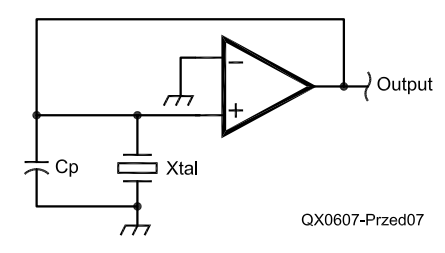
Figure 7 — Oscillator with the crystal in parallel resonance.
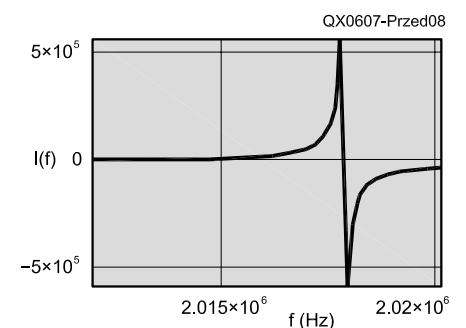
Figure 8 — Reactance versus frequency of the crystal alone.

Figure 9 — Magnitude of the parallel resonant circuit (crystal + C_p).
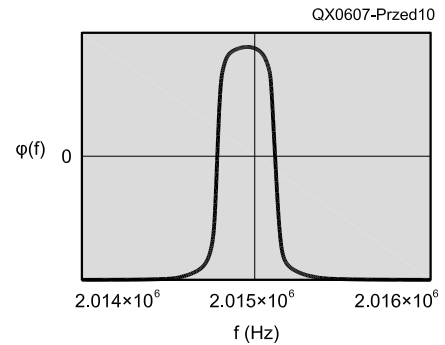
Figure 10 — Phase shift.
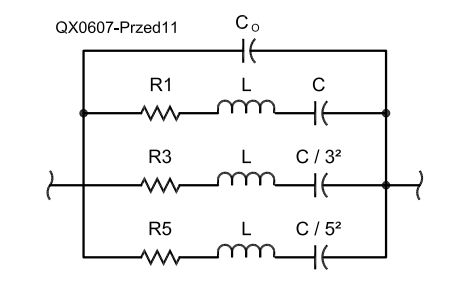
Figure 11 — Equivalent circuit (fundamental, 3rd and 5th overtones only).

also have low noise and should not reduce the overall Q. In very critical applications, a phase-locked-loop circuit can be used to reduce the low-frequency noise contribution of the crystal.

2) Temperature characteristics. The crystal cut, AT, BT, SC, and the angle of the cut determine the frequency versus temperature characteristics. Thus, depending on the application, large temperature range, oven control, and so on, the most desirable cut and angle have to be chosen.

3) Frequency stability. Here, again, a high crystal Q is desirable. Sometimes, even at low frequencies, an overtone crystal is used. Figure 11 shows that a higher Q can be thus obtained. A series resonant circuit is preferred in most cases. The active circuit should have high phase-shift stability.

4) "Pullability." There are applications where it is desirable to change the operating frequency by external means. In this case a lower crystal Q is desired, as shown by Equation 2. If we control the frequency by means of a capacitor in series with the crystal, the maximum frequency control range is approximately equal to:

$$f \times \left(\frac{C}{2(C_o + C_c)} \right) \qquad \text{(Eq 4)}$$

where:

f is the operating frequency and C_c is the tuning capacitor.

5) In addition, if parallel operation is desired, the value of the external tuning capacitor has to be specified. Usually, 32 pF is assumed, unless otherwise specified.

6) Finally, it should be emphasized that *any* crystal can be used in the series or parallel mode. The mode is determined strictly by the active circuit. The main difference is that the parallel resonant frequency is somewhat higher than the series resonant frequency.

Conclusions

As expected, the parallel resonant frequency is higher than the series resonant frequency. Both exhibit very fast phase change versus frequency, indicating a high Q, which contributes to frequency stability and low phase noise. Much more is involved in the actual circuit design to prevent deterioration of the inherently high Q of the crystal.

For high frequency operation, overtone operation is preferred to harmonics. Usually the 3rd or 5th overtone is used. Sometimes the 7th overtone is used, but higher overtones may be unstable and difficult to design.

Bibliography
Frerking, Marvin E., "Crystal Oscillator Design and Temperature Compensation."
Lee, WaiTak P., "The Analogy of Quartz and Coaxial Resonators in an Oscillator Circuit," Rockwell Semiconductor Systems Division.
Przedpelski, A., "Overtone Crystal Oscillators," *RF Design,* September 1991.
Vig, John R., "Quartz Crystal Resonators and Oscillators," US Army Electronics Technology and Devices Laboratory.

Andrzej "Andy" Przedpelski, KC0CWK, was born in Warsaw, Poland in 1927. He received his BS degree from the Massachusetts Institute of Technology, and did graduate work at Northwestern University, and DePaul University.

Andy became interested in radio at an early age. By the time he was 11 he had built a few crystal sets on pieces of wood— real "breadboards!" — and ran a telephone line to his neighbor in an adjacent house. He earned his Technician license in 1998.

During 1946 and 1947, he served in the US Army Air Force as a radar maintenance and field test technician. Since 1948, he has been with ARF Products, Inc, where he has been responsible for development of transmitters, receivers, printed circuit techniques, transistorized frequency meters and direction finders, crystal oscillators, and missile-borne components. He holds numerous patents. Assignments of increasing responsibility led to his position of Principal Engineer, Research and Development Laboratory, from 1963 to 1970, and in late 1970 he was promoted to the position of Vice President, Development. In this capacity he has primarily been a consultant to all engineers in the company, but his duties also included engineering management, technical supervision, proposal writing, theoretical analysis, and circuit and system design.

Since 1992 he has been consulting, as a member of the SHEDD Group, in the design of RF equipment. Recently he has also been a Consulting Editor for RF Design *magazine and is also on the Editorial Review Board for that publication.*

Antenna Options

L. B. Cebik, W4RNL
1434 High Mesa Drive
Knoxville, TN 37938-4443
cebik@cebik.com

A Broadside of Vertical Wires

Arrays of vertical dipoles provide some interesting antenna system options, as we learn from this month's analysis.

We often misunderstand the quest for more gain. The instant picture is a directional beam rotated by a heavy motor at the top of a large steel tower. That is only one version of robbing energy from unused directions and focusing it in the direction of our communications targets. Very often, we can do the required job with a bidirectional antenna. If we can limit our target areas, we can often do away with the tower and the rotator. Wire still works very well, especially in bidirectional arrays.

The set of options that we shall explore in these notes consists entirely of vertical arrays, which are especially apt to the lower HF and upper MF regions. All of the antennas that we shall examine are garden-variety arrays, with many variations in the literature. What may give this treatment a small bit of uniqueness is the fact that we shall level the playing field for our comparisons. To keep everything full size, we shall put every antenna on 40 meters — specifically 7.15 MHz. As well, each antenna will be above average ground; that is, ground with a conductivity of 0.005 S/m and a relative permittivity of 13. Once we examine some basic standards against which to measure array performance, all antennas will use AWG 12 copper wire. As a result, any changes of ground, frequency, and element size will apply roughly equally to all of the antenna types in the group.[1]

[1]Notes appear on page 60.

In addition, we shall bypass the potential long list of monopole arrays and thereby evade questions surrounding ground radial systems and end-fire directional phasing schemes. We shall limit ourselves to antennas based on the vertical dipole and broadside arrays of dipoles that give us bidirectional patterns. The excluded antennas all are good candidates for service in amateur communications, but we have only limited space to conduct a review of some of the possibilities.

A Standard: The ½-λ Vertical Dipole

The fundamental standard against which to measure the performance of all of the other related antennas on our incomplete list is the vertical dipole. Figure 1 provides a snapshot of the antenna's outline and both the elevation and azimuth patterns that emerge under the modeled test conditions. The sample antenna is a 1¼-inch-diameter aluminum center-fed vertical that approximates what we might find

---

Table 1
Modeled 40-Meter Vertical-Dipole-Array Performance Characteristics

Gain dBi[1]	TO Angle[2]	Beamwidth[3]	Feed Point Impedance[4]
Vertical dipole (See Figure 1.)[4, 5, 6]			
0.32	15°	Omni-directional	$75.1 - j0.6$ Ω
Vertical dipoles spaced ½ λ and fed in phase (See Figure 2.)			
4.80	15°	58°	$55.6 + j2.3$ Ω (x 2)
Vertical dipoles spaced ½ λ and fed in phase (See Figure 3.)			
6.25	15°	42°	$55.6 + j2.1$ Ω (center dipole)
			$48.3 + j1.7$ Ω (outside dipoles)

Notes

[1]Except for single dipoles, all antennas are bidirectional broadside to the plane of the vertical wires.

[2]Gain is maximum at the TO angle (or elevation angle of maximum radiation).

[3]Beamwidth is the half-power value for each of the two lobes.

[4]Impedance values are the resistive and reactive series values.

[5]See the indicated figures and the text for the physical description of the individual arrays and their elements.

[6]Test frequency is 7.15 MHz. Ground is average (conductivity = 0.005 S/m; relative permittivity = 13).

[6]These notes apply to all subsequent tables of modeled performance figures.
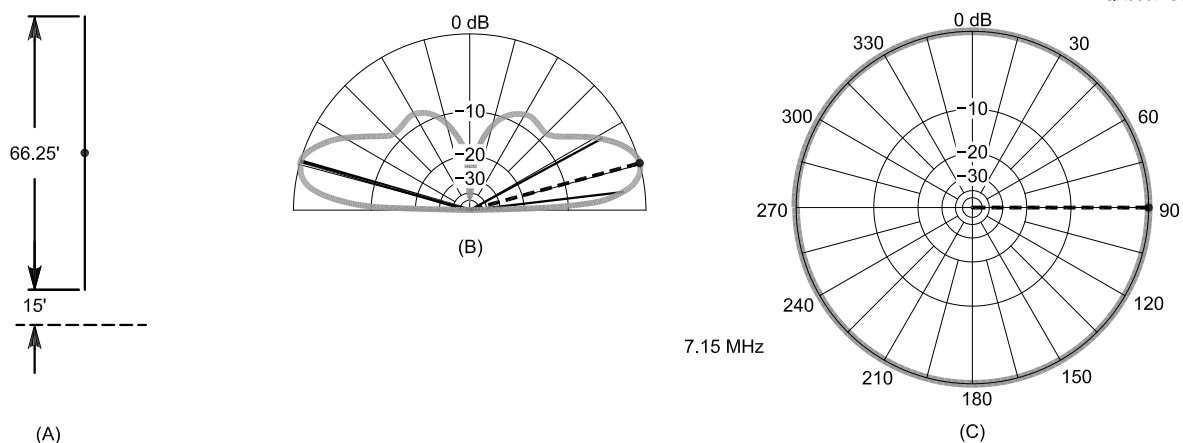
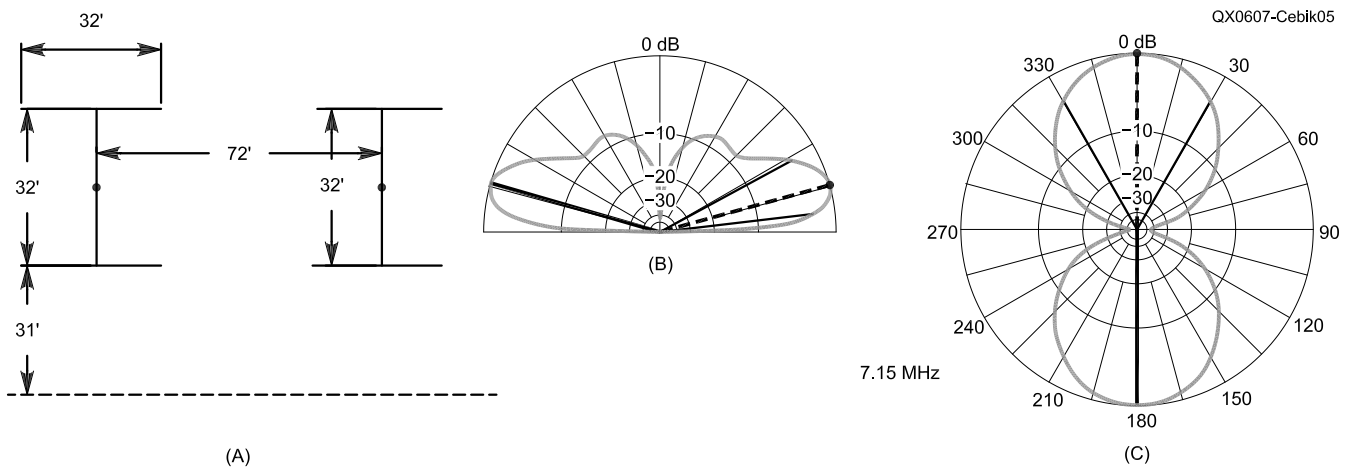Figure 1 — General outline and elevation and azimuth plots for a 40-meter vertical dipole.
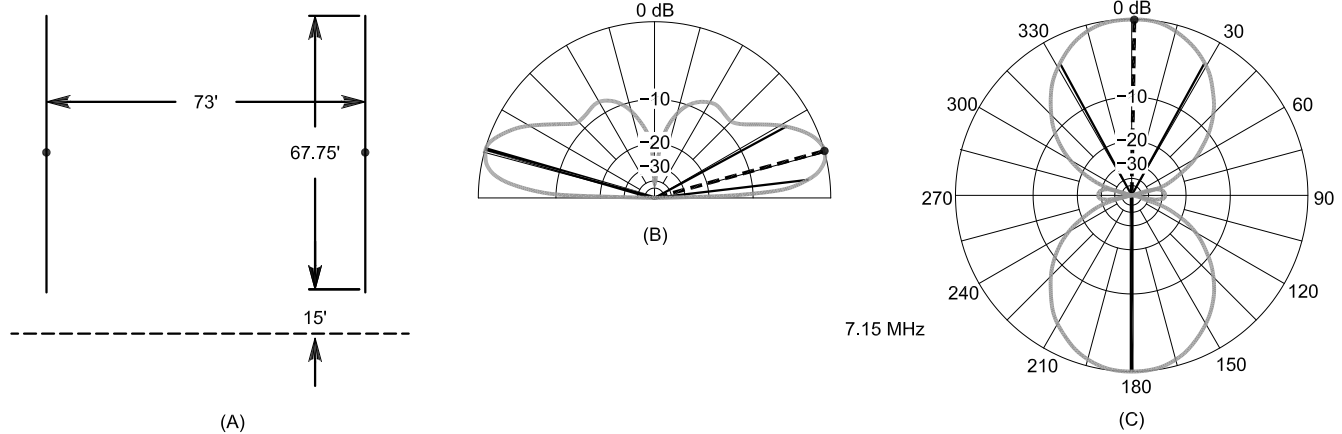
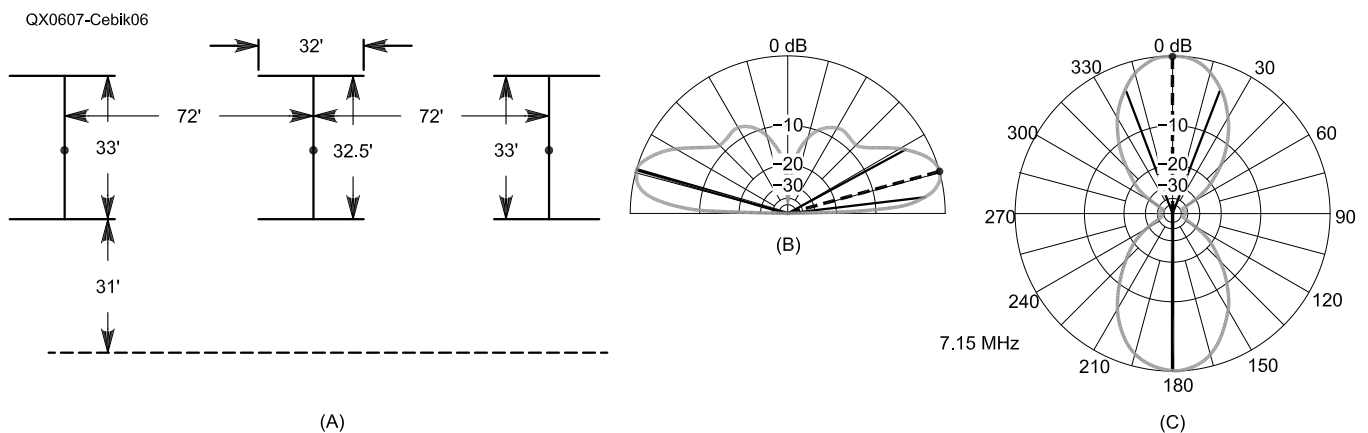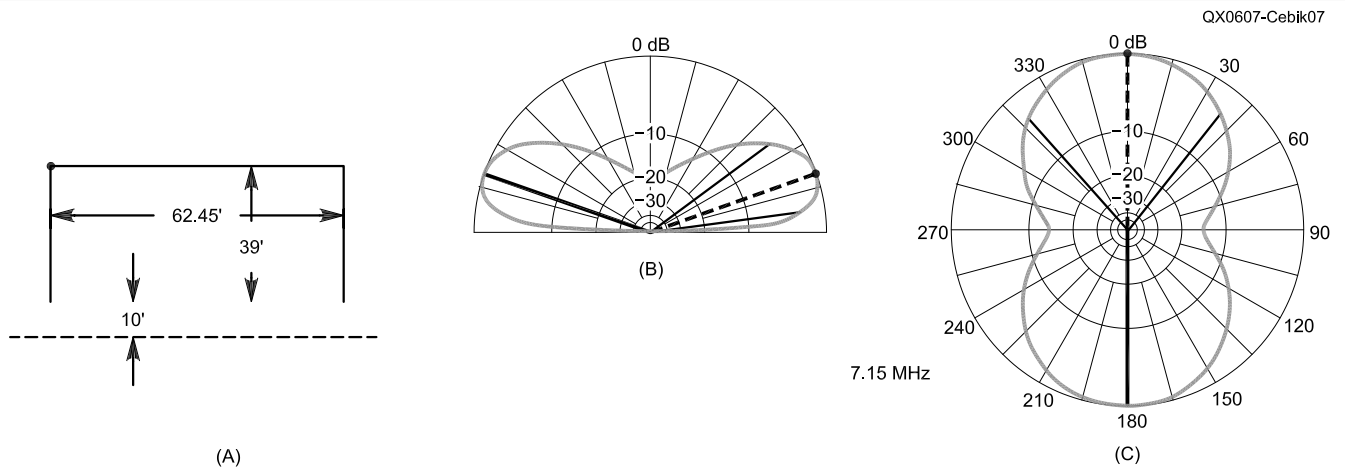Figure 2 — General outline and elevation and azimuth plots for two 40-meter vertical dipoles fed in phase.
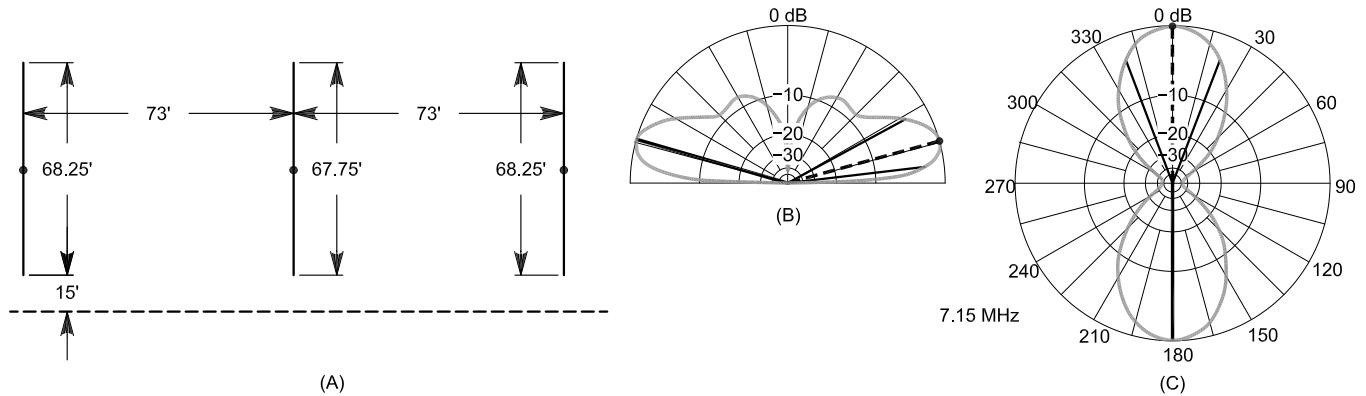
Figure 3 — General outline and elevation and azimuth plots for three 40-meter vertical dipoles, fed in phase.

in actual practice. Real antennas in amateur use might consist of anything from wire suspended from an overhanging limb to large well casings or even light tower sections.

The base of the dipole is about 15 feet above ground. Since a wavelength at 7.15 MHz is just about 140 feet long, you can change the numbers in feet to a fraction of a wavelength. From that point, conversion to metric measures is simple. The importance of the base height shows up in the elevation pattern for the dipole. For any vertical antenna or array that we do not attach directly to ground and a radial system, two performance values interact in opposite directions. As we raise the antenna, the elevation angle of maximum radiation (take-off, or TO angle) decreases slowly and the gain increases. The same increase in height will also enlarge the higher-angle lobe. At a certain height — which varies with each array — all further energy will go into the second or higher elevation lobe. The heights that I have selected may not coincide with what your own

Table 2
Modeled 40-Meter Hatted Vertical-Dipole-Array Performance Characteristics

Gain dBi	TO Angle	Beamwidth	Feed Point Impedance
Hatted vertical dipole (See Figure 4.)[1]			
0.29	16°	omni-directional	36.2 + j2.8 Ω
Hatted vertical dipoles spaced ½ λ and fed in phase (See Figure 5.)[2]			
4.65	16°	60°	27.1 − j1.8 Ω (x 2)
Hatted vertical dipoles spaced ½ λ and fed in phase (See Figure 6.)[3]			
6.11	16°	42°	27.0 − j2.0 Ω (center dipole)
			24.3 − j1.0 Ω (outside dipoles)

Notes: See Notes for Table 1.

physical conditions permit or with your preferences, so I recommend that you reproduce the exercise using your own setup.

My selections rest on several conditions. First, the base height must be at least 10 to 12 feet above ground for safety. The outer

end of any dipole or wire array holds an injurious voltage level. Lowering the base of the vertical dipole would require additional protection measures for people and animals. Second, the higher-angle lobe should be as small as practical to reduce sensitivity to
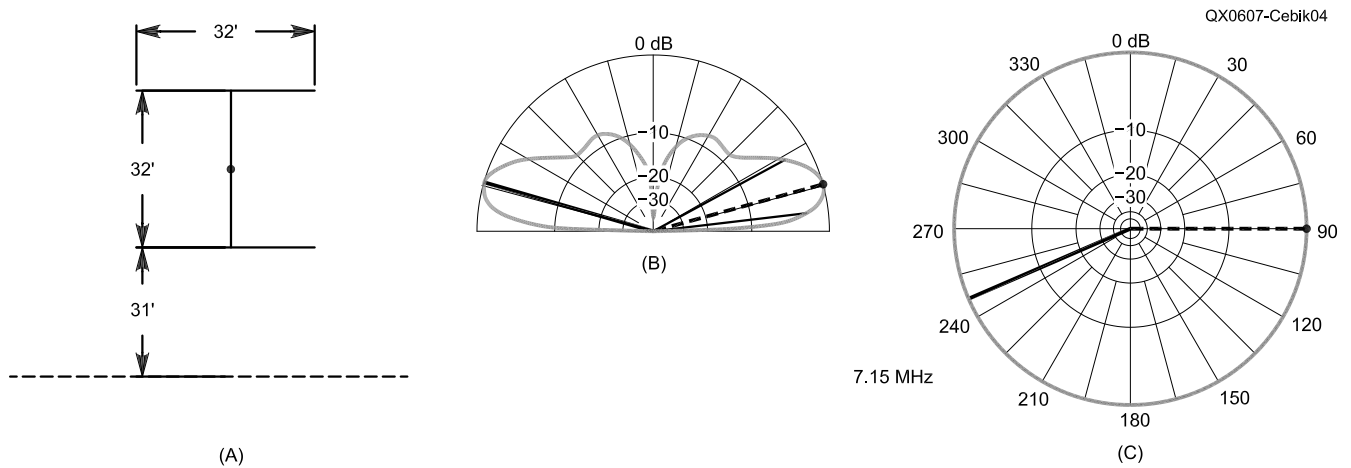
Figure 4 — General outline and elevation and azimuth plots for a 40-meter hatted vertical dipole.

Figure 5 — General outline and elevation and azimuth plots for two 40-meter hatted vertical dipoles, fed in phase.

Figure 6 — General outline and elevation and azimuth plots for three 40-meter hatted vertical dipoles, fed in phase.

higher angle signals. The resulting pattern disqualifies these antennas from NVIS service, but provides quieter background levels for the DX operator. Third, the combination of gain and TO angle must be optimal. This last condition is a judgment call and will vary with the needs and preferences of each operator. Table 1 provides a summary of modeled performance data for my selections for all three vertical dipole antennas and arrays.

The vertical dipole itself is not yet a broadside array, since we have no plane of elements

against which to measure the broadside directions. One simple array consists of two ½-λ vertical dipoles fed in phase. Figure 2 provides the outline and the patterns for the dipole pair. Note that mutual coupling requires that we lengthen the dipoles slightly to obtain resonance. For the test conditions, the bidirectional maximum gain at the 15° TO angle is about 4.5-dB greater than the gain of a single dipole. In exchange, the half-power beamwidth is 58° in each prime direction. Do not underestimate the importance of the beamwidth value in planning a broadside array. Feeding the array is simple in this case, since the dipole feed point values are close to 50 Ω. Equal lengths of coax provide equal feed point currents. If we wish to have 50 Ω at the junction of the two lines, then we must resort to ¾-λ 70 to 75-Ω lines (since the velocity factor of ordinary coax lines will not allow electrical quarter-wavelength lines to reach a common junction).

The azimuth pattern in Figure 2 shows another interesting fact about pairs of vertical dipoles fed in phase. The separation shown is 73 feet, slightly greater than ½ λ.

Table 3
Modeled 40-meter wire vertical array performance characteristics.

Gain dBi[1, 2]	TO Angle	Beamwidth	Feed Point Impedance
Half-square array (See Figure 7.)			
3.46	19°	80°	68.4 − j1.3 Ω
Bobtail curtain (See Figure 8.)			
5.08	18°	53°	52.1 + j2.2 Ω
Single right-angle delta array (Not shown; see Note 3 below.)			
1.90	20°	116°	61.1 − j0.5 Ω
Double right-angle delta array (See Figure 9.)			
3.60	21°	74°	40.6 − j6.5 Ω
Single side-fed rectangle (Not shown; see Note 4 below.)			
3.05	17°	76°	14.6 + j1.4 Ω
Double side-fed rectangle (See Figure 10.)			
4.50	17°	55°	30.9 − j1.0 Ω

Notes
[1]See Notes for Table 1.
[2]All wire arrays use AWG no. 12 (0.0808 inch diameter) copper wire.
[3]The single right-angle delta has a maximum height of 30.4 feet with base of 60.8 feet. The base is 20 feet above ground. The side feed point is 15% up one sloping leg to maximize vertically polarized radiation.
[4]The single rectangle has a maximum height of 12.8 feet, with a horizontal length of 58 feet. The lower wire is 37.3 feet above ground.

Figure 7 — General outline and elevation and azimuth plots for a 40-meter half-square array.
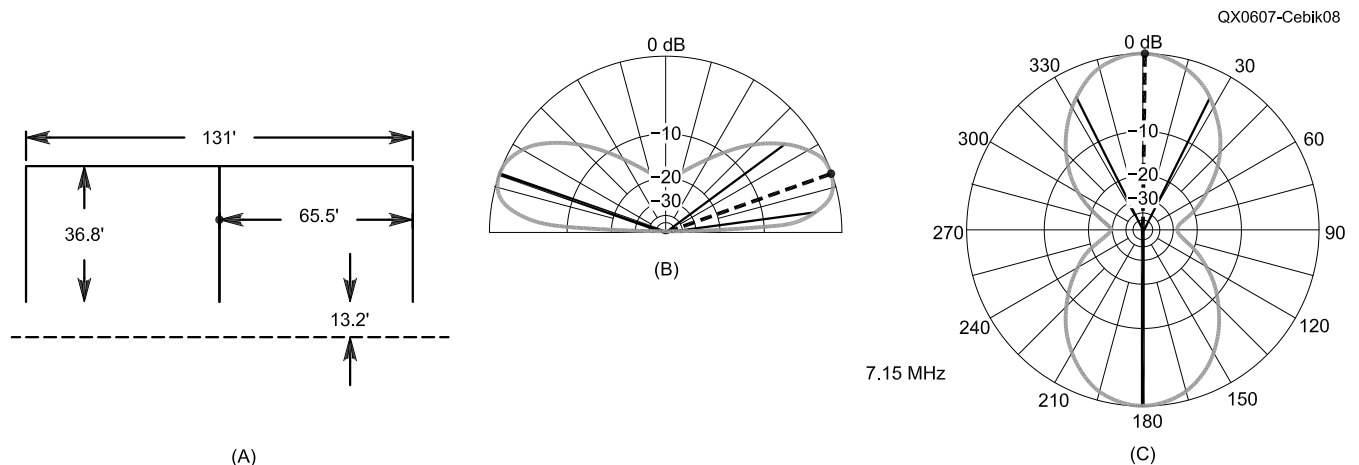
Figure 8 — General outline and elevation and azimuth plots for a 40-meter bobtail curtain.

As we increase the spacing beyond ½ λ, the gain continues to increase, but we find side lobes emerging. At what spacing we terminate the quest for gain because the side lobes are too large is again a judgment call.

Figure 3 shows what we can obtain from three vertical dipoles in phase. We obtain nearly 1.5-dB additional gain, but with a beam that is over 15° narrower. To achieve near resonance at the feed points, we must make the outer dipoles slightly longer than the center dipole. The factor that hinders most antenna builders from implementing a three-dipole array (besides space) is the need for binomial current distribution. To produce the desired pattern, the center dipole must show twice the feed point current as each of the outside dipole feed points. One advantage that accrues to the wire arrays that we shall sample is the use of a single feed point for the entire array.

Hatted Dipoles

Before we look at the wire arrays, we should examine a technique that can overcome to some degree the height requirements of the simple vertical dipole. We can add hats to each end of each dipole, and reduce the vertical length without jeopardizing the performance by very much. The hats can be any symmetrical wire extensions at right angles to the vertical. The extensions permit the antenna to reach a resonant length, but the symmetry of the extensions tends to cancel the horizontal component of the far fields. Figure 4 shows the outline (and patterns) of a 1.25-inch diameter vertical dipole shortened to 32 feet. Each AWG no. 12 hat wire is 16 feet long. The hat in this case takes the form of a T, with the wires in the plane of the dipoles. Hence, we may lash the extension wires to a pair of non-conductive lines at the top and bottom of the dipoles to simplify overall construction. If we raise the feed point level above ground to nearly the same height as the full dipole feed points, we

obtain close to the same TO angle and very close to full gain. We significantly reduce the near-resonant feed point impedance, however. The same considerations would apply to all-wire hatted dipoles, although the vertical dimensions would be somewhat longer. See Table 2 for the modeled performance data for all three hatted arrays.

Figure 5 demonstrates the hatting technique applied to a pair of dipoles fed in phase. We could satisfactorily transform the impedances for a common-junction match to a 50-Ω main feed line using 50-Ω, ¾-λ lines. The gain, TO angle, and beamwidth are comparable to the values for full-size dipoles in the same arrangement. The azimuth pattern shows the consequences of closing the spacing slightly: the side lobes have virtually disappeared. By holding the hat extension wires at a constant 16 feet, we can see the consequence of mutual coupling in the vertical section of the two dipoles. Like the full-size dipole array, the individual elements require lengthening to restore a near-resonant condition.

An array of three hatted dipoles appears in Figure 6. It uses the same spacing (72 feet) as used in the two-dipole array to reduce side lobes. The results once more are very similar to those for full-size dipoles. This array requires binomial feeding, however, just like the full-size counterpart, with the feed point current at the center element reaching twice the value of the feed point currents on the outer elements. I have included the three-dipole arrays not as a recommendation for construction, but instead to provide standards against which we may compare the performance of wire arrays that simulate three-dipole arrays.

Wire Arrays

Almost any antenna handbook contains the names of vertically polarized wire arrays that do not and need not touch the ground: half-squares, bobtail curtains, deltas, and rectangles.

Physical constraints often dictate which among the candidates that we can implement. How we categorize the entire collection depends on the perspective from which we approach them. I once called the entire group "self-contained verticals" or SCVs, since they do not require ground radial systems. We may also view them from the feed point perspective. Each antenna has one feed point but presents vertical wires in phase. Hence, we might call them "self-phasing arrays." By bending and connecting parts of each dipole element so that they touch, we create lines that change the phase of both the voltage and the current, to leave the vertical wire sections essentially in phase with each other at close to optimal spacing.

In our effort to create a level field for comparisons, all of the wire arrays that we shall explore use AWG no. 12 copper wire at 7.15 MHz over average ground. The heights will aim at maximum gain commensurate with a reasonable TO angle, but in no case will the top height exceed 50 feet. Table 3 provides the modeled performance data for several vertical wire arrays.

Perhaps the purest form of SCV is the half-square array. Interestingly, this antenna emerged later than its doubled big brother, the bobtail curtain. (See Woodrow Smith, W6BCX, "Bet My Money on the Bobtail Beam," *CQ*, Mar 1948, pp 21-23 and 92-95, and Ben Vester, K3BC, "The Half Square Antenna," *QST*, Mar, 1974, pp 11-14, for the seminal articles on each antenna.) The half-square is electrically more fundamental, however, corresponding roughly to a pair of vertical dipoles fed in phase. Figure 7 shows the general outline and the plots for a half square at roughly the optimal operating height. We can picture two vertical dipoles. The upper halves of each dipole bend toward each other until they just touch. These touching lines not only complete each dipole, but also form a phasing line between the two corner points. At the center, the
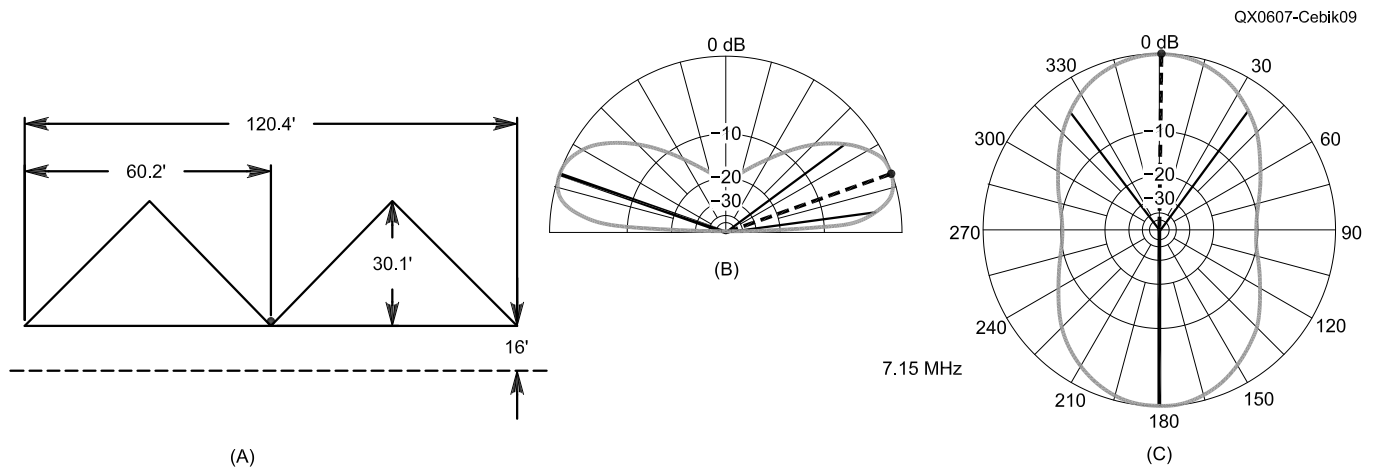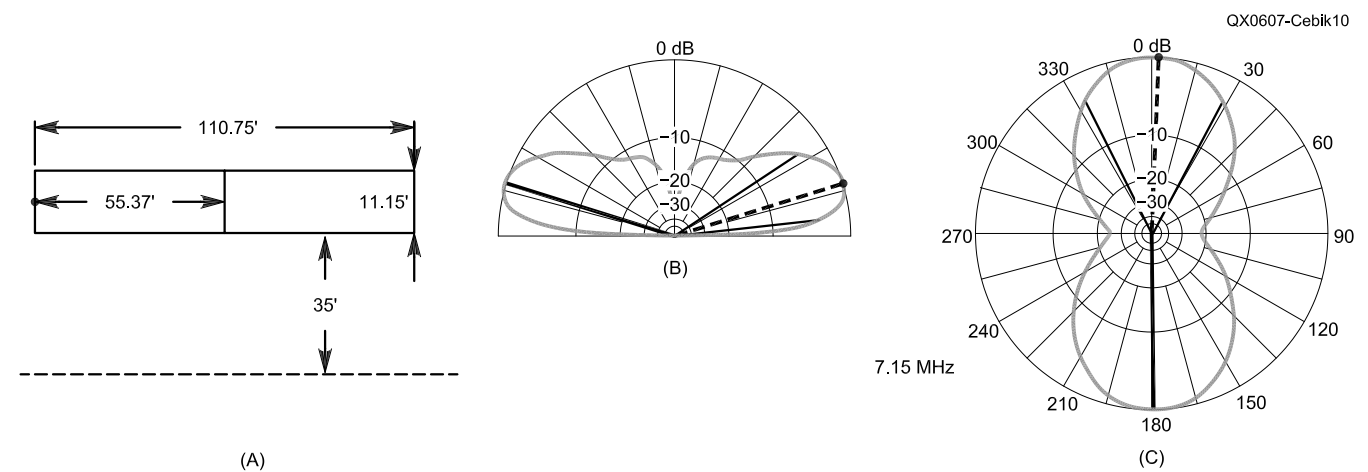

Figure 9 — General outline and elevation and azimuth plots for a 40-meter double-right-angle-delta array.

voltage and current undergo a phase reversal, so the two vertical wires remain in phase. Hence, we need only a single feed point at one of the two upper corners. (We can also extend one vertical leg to the ground and use voltage-feeding techniques.)

For best performance the horizontal wire must be shorter than ½ λ, while the vertical wires are longer than one-half of a resonant vertical dipole. For common copper wire sizes, the ratio of vertical to horizontal is about 5:8. Because the resulting array does not use full ½-λ spacing (about 70 feet) and only half of the virtual dipole contributes to the effective far field, the gain is about a dB shy of the gain of two vertical dipoles fed in phase. At maximum gain, the TO angle is slightly higher, since the bottom ends of the elements are closer to ground. These factors also contribute to the 80° beamwidth of the array.

The bobtail curtain appeared earlier, but is electrically the double of the half-square array. Figure 8 presents the outline and the pat-terns for a roughly optimized bobtail curtain. The dimensions closely correspond to those worked out empirically by SM4CAN. The vertical elements are shorter than those of the half square, but the phase lines in each of the two sections are longer. To obtain a 50-Ω imped-ance, the feed point is about 60% of the way up the center element. Bringing the center element to the ground and using voltage-feeding techniques is very common, however. Still, we lack the radiation from the upper portions of the vertical dipoles and the spacing remains shy of a more ideal ½ λ. Hence, the maximum gain is lower than for three half-wavelength dipoles in phase, and the beamwidth is greater. In fact, the bobtail curtain performance figures more closely resemble the values associated with two half-wavelength dipoles in phase. Nevertheless, the bobtail's single feed point produces a very close approximation of bino-mial feed with its 1-2-1 current magnitude ra-tios at the upper corners of the array.

There are two configurations that many hams use largely to overcome site constraints. One is the delta and the double delta. Table 3 provides data on the single right-angle delta, when we feed it about 15% up one side to maximize the vertical component of the far-field pattern. This compact form that requires only a single upper support falls short of the ideal vertical element spacing and hence shows the lowest gain of any of the wire arrays. Its cousin, the double delta, appears in Figure 9. We feed the array across the upper and lower (or base) wires at the center for vertical polar-ization. Although the feed point impedance is close to ideal, the gain is less than the gain of two vertical dipoles in phase — with a corre-sponding increase in beamwidth.

The side-fed rectangle and double rectangle provide higher gain than single and double deltas largely because the end elements are vertical and closer to ½ λ apart. Table 3 pro-vides data on the single rectangle. Its optimum dimensions represent a compromise between spacing and the required length of the end el-

Figure 10 — General outline and elevation and azimuth plots for a 40-meter double-rectangle array.
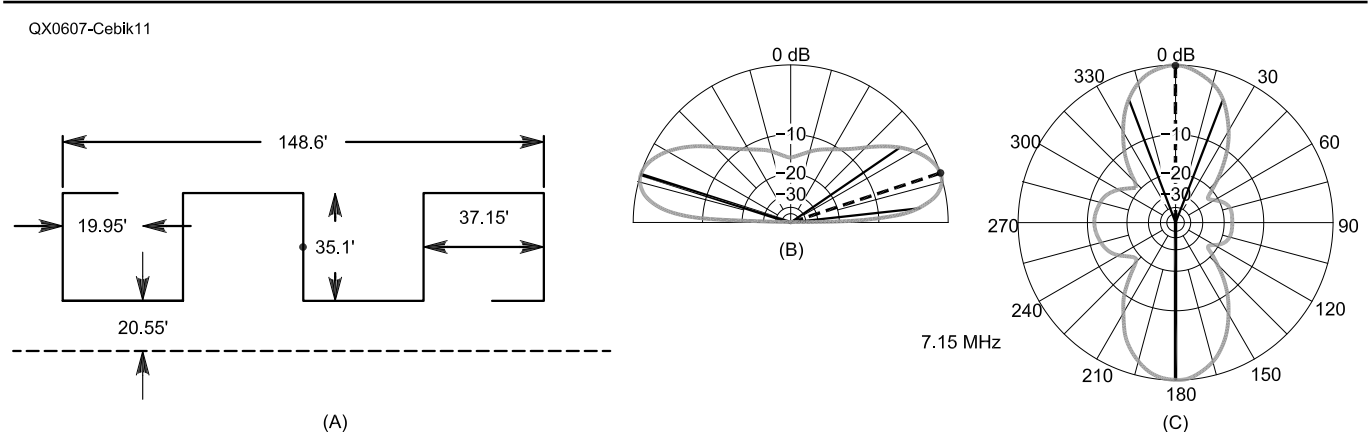
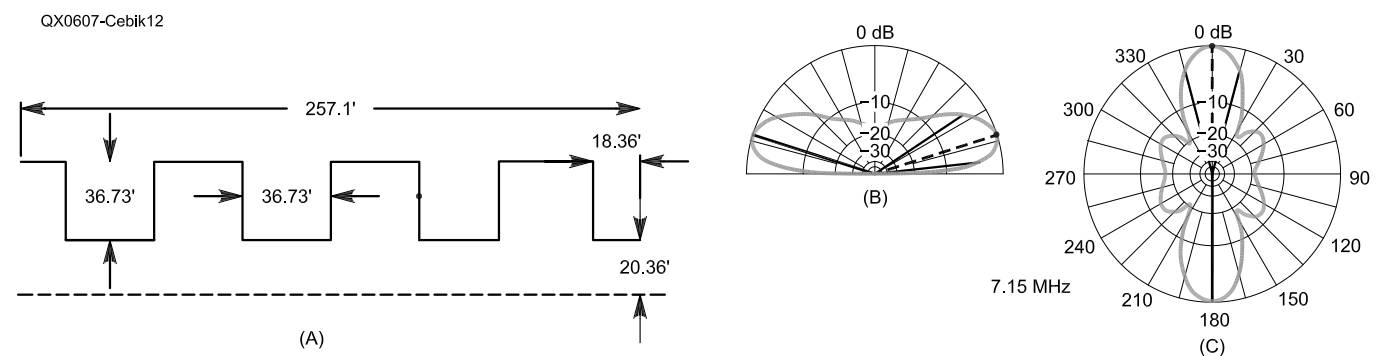Figure 11 — General outline and elevation and azimuth plots for a 40-meter 5-element Bruce array, with in-turned ends.

ements. Element spacing dominates the equation until the vertical end elements become too short to provide a strong far field pattern. The low feed point impedance of a single rectangle tends to reduce the utility of this version of the antenna despite the fact that the spacing of the horizontal wires is relatively small. The double rectangle, shown in Figure 10, is more practical, despite its greater horizontal dimension. The gain and beamwidth approximate the values for two ideal vertical dipoles fed in phase, although the antenna only requires two 50-foot-tall nonconductive end supports. The diagram shows the array fed at the center of one end. The 31-Ω impedance is far more workable than would be the very low impedance at the middle of the center vertical wire, on which we find twice the current magnitude of the end vertical wires.

Of the double arrays, the bobtail curtain and the dual rectangle may be the most popular for those with the required space. The double rectangle is a bit shorter and presents the fewest safety problems. The bobtail curtain, however, has a bit more gain and its loose ends provide a means for voltage-feeding techniques. (Bringing the bobtail center wire to ground and using a tank circuit for impedance matching does not alter the position of the high current point at the corner with the two horizontal phase lines. Nor does it alter the required current magnitudes and phase angles at each

of the array's outer corners.) Half squares and single deltas remain popular among those with more limited space. With a change of feed point, each antenna is adaptable to use as a general all-band wire.

The Bruce Array

The Bruce array or curtain derives its name from Edmond Bruce, one of the legends of antenna developments in the late 1920s and through the 1930s. The array that bears his name is not, according to reports traced to John Kraus, W8JK, the antenna that he would have preferred to bear his name. The preferred antenna is the rhombic (originally named the diamond by Bruce). In fact, the Bruce-type curtain fell out of favor among wire fans in the amateur community until resurrected by Rudy Severns, N6LF, and given prominence in Chapter 8 of the 20th edition of the *ARRL Antenna Book*.[2] We shall briefly examine two forms of the Bruce array: a five-element "innie" and a seven-element "outie." A Bruce array requires $1/8$ λ end sections, and the navel reference refers to how we handle them. Table 4 provides the modeled data for the two sample arrays.

As suggested in Figure 11, a Bruce array consists of a set of parallel vertical dipoles with a vertical height of about ¼ λ. The separation between vertical sections is also about ¼ λ. Hence, each horizontal wire represents the

completion of a vertical dipole. Unlike the half square and the bobtail curtain, the Bruce verticals presume a high current region at the center of the vertical wire sections. Unlike the rectangle, the bottom wire does not return to the position of the top wire, but extends in the opposite direction to connect to the next vertical dipole. The ideal length both vertically and horizontally for wire sections (excluding the shorter end sections) is actually about 1.05 times a quarter wavelength, although the antenna admits of considerable variation. The five-element version represents an attempt to squeeze out the maximum gain from the

Table 5
Reference List of Models and Figures in Which They Are Used

EZNEC *Model*[1]	*Figure*[2]
Vdpl40.ez	1
Vdpl40x2.ez	2
Vdpl40x3.ez	3
Vhat40.ez	4
Vhat40x2.ez	5
Vhat40x3.ez	6
Hs40.ez	7
BC40.ez	8
Dbdelta-40.ez	9
Dbrect40.ez	10
Bruce-5i.ez	11
Bruce-7o.ez	12
Radelta-40.ez	
Rect-40.ez	

Notes
[1]Models for the antennas discussed in this "Antenna Options" column are available in *EZNEC* format at the ARRL Web site at **www.arrl.org/qexfiles**. Look for 7x06_AO.zip.
[2]All azimuth patterns taken at the TO angle indicated on elevation plots. All elevation plots taken at a 90° azimuth angle. No complex or overlapping plots.

Table 4
Modeled 40-meter performance of two versions of the Bruce array.

Gain dBi[1, 2]	TO Angle	Beamwidth	Feed Point Impedance
Five-element array with ends turned inward (See Figure 11.)			
6.09	18°	41°	295.0 − j0.7 Ω
Seven-element array with ends turned outward (See Figure 12.)			
7.36	17°	29°	461.0 − j3.2 Ω

Notes
[1]See Notes for Table 1.
[2]Arrays use AWG no. 12 copper wire.

QX0607-Cebik12

257.1'

36.73'   36.73'   18.36'

20.36'

(A)

0 dB
−10
−20
−30
(B)

0 dB
330   30
300   60
270   90
240   120
210   150
180
−10 −20 −30
(C)

7.15 MHz

Figure 12 — General outline and elevation and azimuth plots for a 40-meter 7-element Bruce array with out-turned ends.

number of elements, which required slightly more spacing horizontally, and slightly less vertical length. The benefits are very small for the exercise, however.

The most interesting fact about the five-element version is the inward turn of the end sections. The resulting array is 148 feet long, only slightly longer than longest of the doubled wire arrays. Using a centered feed point on the center wire (other feed point positions are possible), the array produces just over 6 dBi gain, close to the maximum that we can encourage from three vertical dipoles fed in phase. The modeled feed point impedance is close to 300 Ω, a handy value for available transmission lines. The one major difference from other arrays occurs in the azimuth pattern. It shows side lobes that cannot be eliminated. They result from spacing vertical elements at ¼-λ intervals. The differential in the side lobe sizes results from

one end section occurring at the array top and the other existing at the array bottom.

The alternative form of the Bruce array — shown mostly in college texts such as Kraus' *Antennas* — points the end sections outward. The seven-element version of the array appears in Figure 12 and requires the horizontal span that the inward turned version would need for eight sections. Using seven elements allows a central wire for the feed point. It uses a more conventional set of element dimensions, with equal vertical and horizontal lengths. Nevertheless, adding two more sections has two consequences. First, the impedance is close to 460 Ω, another convenient impedance for available transmission lines. Second, the larger version shows increased gain over the smaller version. The increase is about 1.25 dB, with a commensurate shrinking of the beamwidth. The progression of gain-versus-beamwidth ratios

provides a glimpse at the original use for curtain-type arrays. In the 1930s, point-to-point communications with specific cities across the ocean comprised a significant set of commercial enterprises. Hence, both Bell Labs and RCA gave high emphasis to more directive wire arrays. Unless one is very favorably positioned between two major target areas 180° apart, it is possible to set up a wire array with too much gain and too little beamwidth. The azimuth pattern for the seven-element Bruce array shows the diminishing beamwidth when compared to other azimuth patterns in the series.

An Interim Conclusion

We have not examined every vertically polarized wire array that we might use on 40 meters, indeed, not even every wire array based on the vertical dipole. We have subjected enough types to comparable modeling conditions, however, to allow some evaluation of performance potential versus mechanical requirements for installation. All of our samples except the original full-size vertical dipoles might easily suspend from sturdy ropes between nonconductive 60-foot end support posts, trees, or even surplus telephone poles. Scaling the arrays to 80/75-meter size would likely require lower bottom wires or element ends in addition to taller posts. The results would include increased TO angles and slightly lower maximum gain levels. Even so, vertically polarized wire arrays offer a large number of options for bidirectional lower-band antennas.

These notes have omitted the many bidirectional horizontal arrays of use to radio amateurs. I shall partially fill the void next time. In the meantime, be sure to obtain for your library not only a current edition of *The ARRL Antenna Book*, but as well a copy of ON4UN's compendious *Low-Band DXing*.[3]

Notes
[1]Models for the antennas discussed in this "Antenna Options" column are available in *EZNEC* format at the ARRL Web site at **www.arrl.org/qexfiles**. Look for 7x06_AO.zip.
[2]*The ARRL Antenna Book*, 20th Edition, is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 9043. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.
[3]J. Devoldere, ON4UN, *Low Band Dxing,* 3rd Edition, is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 9140. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.

Letters to the Editor

An Analysis of Stress in Guy-Wire Systems (Mar/Apr 2006)

Doug,

On page 42 of my article, the pressure on a round cross-section is defined as $P = 0.0025 \, V^2$. I shouldn't have stated that $FW = \text{Area} \times \text{Pressure} \times \text{Coefficient of Drag} = WW \times RD \times P \times Cd$, since the coefficient of drag (0.0025) is already incorporated into the pressure value. In fact, that is true wherever the pressure appears in an equation. The attendant computer program is correct.

In an equation on page 43, one character was omitted. The equation is $dT = y \times dF = y \times P \times dA = \text{lever arm} \times \text{pressure} \times \text{width} \times y$ (the last term should be dy rather than y). I missed that on the proof copy.
— *73, Bill Rynone, Ph.D., P.E., PO Box 4445, Annapolis, MD, 21403*

RF Power Amplifier Output Impedance Revisited (Jan/Feb 2005; *Letters*, Mar/Apr 2006)

Hello Doug,

This letter is in response to your challenge to the readers of Letters to the Editor in the March/April issue of *QEX*, and to the two contributors in that issue debating the topic of RF power amplifier output impedance for solid-state PAs, that certain assertions previously advanced on these pages remain unchallenged, and that further experiments are needed. I would like to remind you that an overwhelming series of experiments to convince readers that when a PA tuned for maximum power output, and operating within the design recommendations by the manufacturer of the tubes used, is indeed conjugately matched to its load.[1] My response refers to your comments on the load variation method to measure the output impedance of a power amplifier, refuted by Warren Bruene, W5OLY, "Letters to the Editor", Jan/Feb 2001 *QEX*, pp 59-61.

I do not intend to rebut that letter. My purpose is to convince you that Mr. Bruene is wrong, based on experiments previously reported.

Tom Rauch, W8JI, (a co-author of the referenced paper) improved the test procedure devised by W5OLY, which is involved with feeding a small reverse generator signal back into an operating amplifier via a high-power attenuator, and measuring the reverse generator's voltage level along a 50-ohm transmission line. This test once again agreed with the results I obtained using the test setup identical to Mr. Bruene's, but the new test's ability to determine the direction of change resulted in conclusions very much contrary to Bruene's earlier measurements.

Mr. Rauch's measurements revealed that, for some 14 amplifiers of widely varying types, maximum efficiency could be obtained by tuning the output network while solely observing the reverse mismatch change! As the tank network was adjusted to present a 50-Ω load to the reverse power generator (reverse generator voltage equal at every point along the 50-Ω line), maximum efficiency and output power was obtained. As a matter of record, Rauch noted it was much more difficult to obtain optimum efficiency using the meters on the amplifier and the power output indicator than it was by watching the mismatch for the reverse power generator (RPG).

Amplifier output impedance (referenced to the output terminals of the PA) is certainly non-dissipative; power generated is available for transfer. Assuming a low-loss transmission line, the impedance of the transmission line is a non-dissipative impedance. The antenna itself has a measured (or calculated) input impedance, which for efficient antennas is a non-dissipative impedance. Power is not absorbed by the resistive component of this impedance, power input to the antenna is transferred to the propagation medium. Finally, since the input impedance of antenna systems measured at the input to the transmission line feeding the antenna is generally not a resistive impedance equal to 50 Ω, an antenna system tuning unit (ASTU) is used, the purpose of which is to provide a conjugate at the output terminals of the ASTU, and hence a conjugate match referenced to the input terminals (the transmitter side of the ASTU).
— *John S. (Jack) Belrose, VE2CV, ARRL Technical Advisor,* **john.belrose@crc.ca**

Hi Jack,

Thanks for your letter. We had to shorten it a bit so we could focus on two fundamental assertions you mention that appear to us to be mutually incompatible.

You've consistently stated that maximum power transfer occurs when a conjugate match is achieved. Yet, you indicate that under the conditions you claim to constitute such a match, no power is dissipated at the tube end of a network from the reverse power injected during the Bruene experiment (non-dissipative resistance), even though Mr. Rauch measures the s22 of the amplifier to be 50 + *j*0 Ω.

In that case, you imply on the one hand that *no* power transfer is occurring from load to source and on the other, that no reflections occur anywhere. We just don't see how you can have it both ways.

Were amplifier output impedance completely non-dissipative, you would not measure an s22 of 50 + *j*0 Ω. You would instead measure a pure reactance. If you say that the s22 isn't the same as the amplifier output impedance, then you don't have a conjugate match by definition.
— *Doug Smith, KF6DX,* QEX *Editor,* **kf6dx@arrl.org**

In Search of New Receiver Performance Paradigms (*Empirical Outlook*, May/June 2006)

Dear Doug:

Your plaints concerning measured IMD shortcomings are well understood if not widely recognized. More specifically:

The voltage gain of nonlinear, black-box components (such as receivers, transmitters, A/D converters, etc) can be expanded in a power series when input level does not cause significant change in component operating point. For typical low-distortion components, that power series can often be truncated, retaining terms only up to the third order. Straightforward trigonometric expansions of two-tone response then yield the often useful concepts of second- and third-order "intercept points." These points permit a rapid estimation of useful black-box dynamic range. It is quite easy to show that these estimates are seriously in error when:

1) More than two sinusoids are applied to the input (complex waveforms, multiple interferers, etc).

2) Truncated terms above third order are significant (such as A/D converters, class-C transmitters, etc).

3) Black-box operating point is a function of input level.

In those instances, detailed and often messy calculations are required, based upon both the actual voltage-gain function and the phase relationships among the multiple input sinusoids. I know of no simple ways of overcoming these inherent drawbacks to the use of IMD in characterizing black-box performance.
— *Neal Eddy,* **fneddy@charter.net**

Hi Neal,

Thanks for your comments. My main point, of course, was that we continue to report figures that don't comply with the defining equation. Something has to give.
— *Doug Smith, KF6DX,* QEX *Editor,* **kf6dx@arrl.org**

[1]J. S. Belrose, W. Maxwell and C. T. Rauch, "Source Impedance of HF Tuned Power Amplifiers and the Conjugate Match," Fall 1997 *Communications Quarterly*, pp 25 - 40.

Measuring Height With a Poor Man's Gizmo (*Tech Notes*, May/June 2006)

Doug:

Last week, I mused (honestly!) about the many times I have heard hams state the heights of their antennas. All of them apparently used the elusive "eyeball algorithm" to establish their measurements. I've never heard anyone say that the height was *measured* with an instrument of known accuracy.

Kudos to William Rynone and *QEX* for the "Poor Man's Gizmo." Print and sell a bunch of reprints and publicize the gizmo in *QST* and on the ARRL Web site. Finally, reflect on the wisdom of Henry St. John (1716): "Truth lies within a little and certain compass, but error is immense."

— *73 de Jim Olsen, Jr, W3KMM,* **w3kmn@aol.com**

Uniform Current Loop Radiators (May/June 2006)

Editor,

NP4B has written an interesting article. What a novel idea to segment a piece of twin-lead like this — very clever. I don't think the theoretical explanation is correct though.

First, the sinusoidal distribution of current along a conventional wire antenna is not due to wire inductance as the article says. It is due to propagation delay, and reflection from the ends of the wire in the case of a dipole. The current distribution is a standing wave along the antenna, caused by the interaction of forward and reverse waves. That is covered in *The ARRL Antenna Book* and elsewhere.

Also, the model for the segmented line shown in Figure 2B doesn't seem to be correct. It appears that the inductors represent the alternating wire segments and the capacitors represent the overlap between the segments. The overlap is almost the entire portion of the line, however, and not only is there capacitance in this region but the wires are magnetically coupled, too. For the entire overlap, in fact, the structure remains the original transmission line with its original distributed capacitive and inductive coupling between the wires.

I think the model for the segmented line is more complex than the author has indicated. The results are interesting and obviously the antenna works, but I don't think the explanation is correct.

— *Gerrit Barrere, KJ7KV,* **gerrit@exality.com**

Dear Doug,

I accept the criticism of KJ7KV on my simplistic explanation as due to wire inductance alone. The model of Figure 2B is a lumped model of a distributed system. The complete distributed system has been modeled on *EZNEC* as in Figure 1 of this letter. Note the (essentially) triangular distribution of current on each wire segment; summing the current in adjacent wires yields a net uniform current, as proposed and realized.

While KJ7KV is technically correct in that the wire segments are also magnetically coupled, this coupling is lower than the electrical coupling by a factor of μ_0 (approximately) and may be ignored.

I look forward to further discussion in *QEX* regarding this antenna.

By the way, my article contained an error in Figure 9, the SWR plot. The vertical scale should be corrected to run from 1.0 to 6.0.

Also, several readers wrote to ask how the feed line is connected and where. Figure 2 in this letter shows how a resonant section is split in two to accept the feed line.

— *73, Bob Zimmerman, NP4B,* **zimmo2@juno.com**

Dual Directional Wattmeters (May/June 2006)

Doug:

There were a few minor errors in my DDW article, and I would like to issue a correction or clarification in an upcoming issue.

In Figure 6B the wattmeter readings should have been 43.5 W and 143.5 W for P_{REF} and P_{FWRD}, and in Figure 6C they should read 5.5 W and 105.5 W.

In Figures A1, A2, and A3, I used an alternate notation for the forward and reverse voltage with V+ used for V_{FWD} and V− for V_{REV}.

— *73, Eric von Valtier, K8LV,* **EVonvaltie@aol.com**

Figure 1 — Detailed *EZNEC* modeling of the uniform current loop antenna with closely spaced overlapping wire segments. For ease of modeling, eight sections are used in this model. The driving source is in the center (current maximum) of element #1.
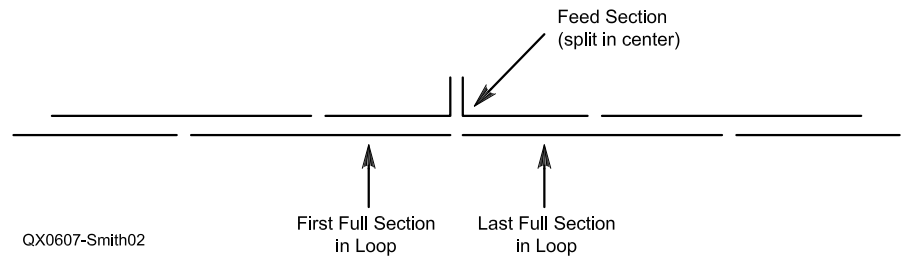
Figure 2 — This diagram shows how a resonant section of the loop is split in the center to accept the feed line.

Out of the Box

Raymond Mack, W5IFS
17060 Conway Springs Ct
Austin, TX 78717-2989

Digital Signal Processing Technology
Doug Smith, KF6DX

Doug Smith is our *QEX* editor, and he insisted I do a review of other DSP (digital signal processing) and software defined radio books to compare with his, before writing this review. The goal was to provide unbiased coverage. Unfortunately, there just isn't a lot of information available outside of college text books on either subject. I did a search of Amazon.com and made trips to Barnes and Noble and Walden Books in several major cities throughout the US. My search proved fruitless.

There are two major things to recommend Doug's book: it is well written, and you can actually purchase it without searching for days and spending an entire paycheck.

Doug covers a broad range of subjects with just enough math to help you understand. This is a book to give you an introduction to DSP and give you some ideas of what you can accomplish. It is not aimed at giving you direct help in producing a digital radio.

He starts with some history and examples of DSP in areas other than radio. Then he moves to a description of sampling, digital number systems, and analog to digital converters. The chapter on digital filters briefly covers all the classical filter types. Doug spends several chapters illustrating radio uses for DSP, including modulation systems, speech coding (for digital transmission of speech), signal synthesis (for transmitter and receiver oscillators), and noise reduction. One of the final chapters covers commercially available silicon and the next covers commercially available (and also free) software for DSP applications. The last chapter and the appendices briefly cover advanced topics.
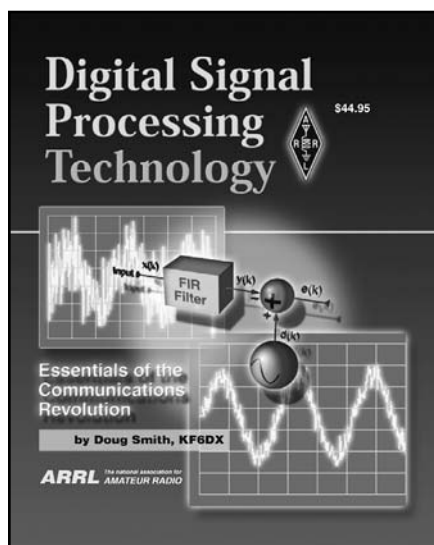
The book was first published in 2001, so the chapters on commercially available resources are a bit dated. This area of electronics changes very rapidly. Fortunately, the bulk of the material in the book is basic and will never be dated.

Digital Signal Processing Technology, Doug Smith, KF6DX, ISBN: 0-87259-819-5, $44.95. *Digital Signal Processing Technology* is available from your ARRL dealer or the ARRL Bookstore, ARRL order no. 8195. Telephone 860-594-0355 or toll free in the US 888-277-5289; **www.arrl.org/shop/**; **pubsales@arrl.org**.

QEX

Celebrate 25 Years of Digital Innovation!

Come to Tucson, Arizona for the 2006 TAPR/ARRL Digital Communications Conference, September 15-17 at the Clarion Airport Hotel. Join with your fellow hams to celebrate the 25th anniversary of the founding of Tucson Amateur Packet Radio, the group that pioneered amateur digital innovation.

Learn about new digital modes, see live demonstrations and share in the camaraderie that is unique to Amateur Radio. Bring your family, too. Tucson is a desert oasis of beauty and fun!

See the Digital Communications Conference site on the Web at **www.tapr.org/dcc/**, or call TAPR at **972-671-8277** to make your reservations today.

QEX- **Jul/Aug 2006 63**

In the next issue of QEX

For many reasons, an accurate frequency standard for radio communications and measurement is desirable. In our next issue, Bertrand Zauhar, VE2ZAZ, presents "A Simplified GPS-Derived Frequency Standard."

Bertrand's unit allows you to supply a very accurate 10-MHz signal to the reference inputs of your transceivers and test equipment.

The design uses the 1-Hz pulsed output used by many GPS units to frequency-lock a 10-MHz VCXO. Claimed accuracy is 1 part in 10^{10} — or the equivalent ratio of 1 Hz to a 10-GHz signal. Full schematics are provided and firmware is available for free download. This is your chance to add a precise and traceable frequency standard to your lab or shack for under $200, including the GPS receiver. Don't miss it! **QEX**

NEW Publications from ARRL

Basic Radio— Understanding the Key Building Blocks

by Joel Hallas, W1ZR
ARRL Order No. 9558 .... **$29.95***
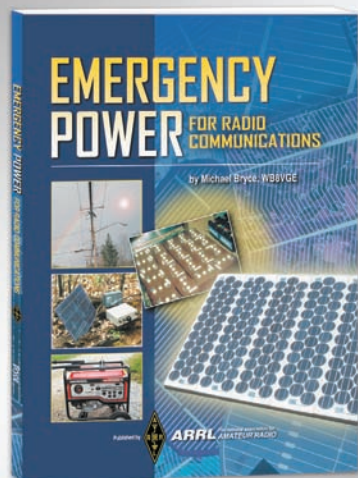*shipping: $8 US (ground)/$13 International

Basic Radio brings the magic of radio to life!

FINALLY— an introduction to radio FOR EVERYONE! — what it does and how it does it.

Basic Radio reveals the key building blocks of radio: **receivers**; **transmitters**; **antennas**; **propagation**; **radionavigation**; and **radiolocation**. This book includes simple, build-it-yourself projects to turn theory into practice—helping reinforce key subject matter.

WHO NEEDS THIS BOOK?

Basic Radio builds upon knowledge of elementary electronic concepts as presented in ARRL's *Understanding Basic Electronics* or a similar course. This book will provide the foundation in radio theory and practice necessary for anyone undertaking more advanced topics such as those presented in *The ARRL Handbook for Radio Communications*.

Includes Build-it-Yourself Radio Projects!

Emergency Power for Radio Communications

by Michael Bryce, WB8VGE
ARRL Order No. 9531........**$19.95**
*shipping: $7 US (ground)/$12 International

With this comprehensive guide you can explore the various means of electric power generation—from charging batteries, to keeping the lights on. This book covers the foundation of any communications installation—the power source. Use this book to plan ways to stay on the air when weather or other reasons cause a short-term or long-term power outage. **When all else fails…how will you communicate?** Find ways to reach beyond the commercial power grid. Identify methods for alternative power genera-tion that will work best in your particular situation, perhaps taking advantage of possibilities already on hand.

Contents

- Keeping the Signals on the Air
- Hey, I Am In The Dark: Keeping The Lights On In The Ham Shack With Emergency Power
- Solar Power
- Charge Controllers for PV Systems
- Generators: Gas, Wind and Water
- Load Sizing
- Holding Your Volts: Battery Systems and Storage
- Systems for Emergency Power
- Inverters
- Station Instrumentation
- Safety
- Emergency Practices

Tools for…
Emergency or Backup Power!
Energy Independence!
Portable Energy!