DEMAND THE BEST

DJ-580T *Best Performance*

The <u>Right</u> Features: Alinco's DJ-580T offers powerful user friendly features that make operating easier. Many new radios available today are so complex that an instruction manual needs to be carried with the radio. We have designed the DJ-580T to be user-friendly, but also sophisticated: It has the features you need, without any useless bells and whistles. This radio can receive 130-174 MHz, 420-474 MHz, and can be modified to receive AM from 118-142 MHz. Other outstanding features include cross-band repeat function, full duplex operation, backlit keypad, 8 scanning modes, CTCSS encode / decode, and DSQ for private paging.

The DJ-580T will deliver a solid 5 Watts of power with the optional 12V battery, and our EMS-8Z Remote Speaker Microphone offers remote control of memory channels and scanning features.

DJ-F1T / DJ-F1TH *Best of Both Worlds*

Small size combined with durability? It sounds like Alinco is breaking the rules again.

Most of the 'micro' handhelds that we've checked out seem to suffer from the same problems: Weak audio, flimsy construction, difficult to use keypads, and poor performance. The DJ-F1T overcomes these problems, offering reliability and performance that surpasses any handheld radio available today. Our unique 'clam shell' battery design allows for a full size keypad for convenient use, and a large speaker for outstanding audio quality. All this is packed into a cast aluminum frame and tough poly-plastic case that has set a new industry standard for durability.

Standard features include: 40 memory channels (odd splits on all), 3 output power levels, DSQ code squelch and private paging, illuminated DTMF keypad, CTCSS encode, Automatic Power Off, and 8 scan modes. The DJ-F1T will receive 130-174 MHz and can be modified to receive AM from 118-142 Mhz.

Better Products, Better Service.
See for yourself why people are coming to ALINCO.

ALINCO ELECTRONICS INC.

438 Amapola Ave., Ste. 130
Torrance, CA 90501
Phone: (310) 618 - 8616
Fax: (310) 618 - 8758

Other Great Products from ALINCO

DJ-G1T

DJ-180T

DR-600T

DR-130T

COMMUNICATIONS QUARTERLY

THE JOURNAL OF
COMMUNICATIONS
TECHNOLOGY

CONTENTS

Littlefield, page 8

Kaplan, page 13

Cover Photo: Global positioning systems (GPS) are becoming increasingly popular as a means of determining one's location anywhere on the globe. Today's new handheld units (great for grid square enthusiasts) make this new technology more affordable. In our cover shot, Bryan Bergeron presents his impression of the new technology. Elliot Kaplan's article on GPS begins on page 13.

EDITORIAL

Don't Shoot Me, I'm The Messenger!

Remember all the flap a few years ago, when an "illegal" message was forwarded without review through a number of automatically controlled packet radio stations? The message provided information about a 900 number interested parties could call to support efforts to keep the United States out of a war in the Middle East. It was initiated by one station and was automatically retransmitted by a number of additional packet stations, before a ham in the Norfolk, Virginia area brought it to the attention of the local FCC office. As the stations involved in this incident faced FCC disciplinary measures for using amateur radio to facilitate the business activities of the *Coalition to Stop U.S. Intervention in the Middle East*, the whole packet radio message-forwarding system came under scrutiny.

At the time, many of us wondered if the FCC's actions would effectively eliminate automatic packet systems and sharply curtail the activities of packet BBSs, altogether. I think we all understood the limitations placed on us in Part 97.113(a) of the FCC rules in regards to business activities on the ham bands; however, given the amount of mail some packet nodes handle each month, we realized it would be a Herculean task for the station sysop to review every message that passed through the system.

Fortunately, the FCC has given packet bulletin board systems and their operators a bit of a reprieve. In April, the Commission released its "Report and Order in PR Docket 93-85" concerning messages relayed by high-speed amateur networks. The new rules took effect on June 1, 1994, and established what the FCC calls "a compliance policy for amateur stations participating in automatic message-forwarding systems."

Under the newly implemented regulations, the majority of station operators in automatic message-forwarding networks will not be responsible for the content of the messages relayed by their stations. This is a change from the previous rules, which stated that all licensees in a chain of forwarding stations were responsible for the content of messages passed through the system. Message *originators* are still responsible for the content of the traffic

they send, and the first forwarding stations are responsible either for the content of the message or for verifying the originator's identity. Of course, any operator in the chain is responsible for discontinuing any communications that violate the rules—as soon as such messages are discovered. In an adjunct to the ruling on automatic packet systems, the FCC determined that, *under the new rules, repeater control operators* will also be relieved of responsibility for inadvertent retransmission of communications that violate the rules.

These rule changes seem to lay the blame for unauthorized transmissions squarely on the shoulders of the perpetrator—a welcome turn of events in these days of "not me, it's not my fault." But do they? I'm still troubled by the fact that the operator of the first forwarding station must verify the identity of the sending station on behalf of the system, or accept accountability for any violations contained in any message he retransmits. Of course, the option is the operator's either check out the identity of the originating station, or accept the punishment should a message in violation of the rules get through. However, while this ruling eliminates the imposition of multiple fines for violations as the message moves along the network, it continues to defeat the purpose of the automatically controlled packet system and punish those not responsible for breaking the rules—those who have come to enjoy the rapid transmission of data via the high-speed packet link. Furthermore, it deflects full responsibility for any illegal messages from the originator, and forces an essentially innocent person to take part of the blame.

In fact, if you think about it, the FCC ruling is tantamount to saying that automatic traffic is now okay *as long as the mailman reads the mail before he takes it to the post office.* I suppose this is better than requiring everyone to read it while enroute. However, the need to sort through all the messages before routing them places the responsibility for making sensitive value judgments regarding content upon people whose primary interest is technical (that is, making sure the system works).

Now, it is possible for packet control opera-

Data Controllers Put You In Control

PK-900

- Dual simultaneous ports.
- Automatic signal idenfication.
- 9600 bps modem option available.
- 8-pole Chebyshev bandpass filter.
- Dual-port Gateway: Packet to AMTOR, Packet to PACTOR, and Packet to Packet.
- Twenty software selectable modems.
- 17K Packet, PACTOR, and AMTOR MailDrop.
- 16-level gray scale fax with opt. software.
- Same modes as the PK-232MBX.
- Full PACTOR Memory ARQ.

DSP-2232
DSP-1232

- DSP-1232 has two switchable ports.
- DSP-2232 has two simultaneous ports.
- Same modes as the PK-232MBX.
- Includes modems for satellite use.
- Automatic signal identification.
- 17K Packet, PACTOR, and AMTOR MailDrop.
- DSP-2232 has dual-port Gateway: Packet to AMTOR, Packet to PACTOR, & Packet to Packet.
- DSP-2232 features front panel LCD.
- Up/Down Doppler shift for PSK modems.
- 9600 bps modem included.

PK-232MBX

- Designed for multi-mode operation.
- Internal 1200 bps VHF modem.
- Automatic signal identification.
- 18K Packet, PACTOR, and AMTOR MailDrop.
- 2400 bps modem option.
- 8-pole Chebyshev bandpass filter.
- Modes: Morse code, Baudot, RTTY, ASCII, AMTOR/SITOR, PACTOR, VHF & HF packet, B&W fax receive/transmit, NAVTEX/AMTEX, and ARRL Information Services.
- Gateway as a node.

PK-12
PK-96

- PK-96 is a 9600 bps packet-only controller with 9600 bps K9NG and G3RUH compatible direct frequency modulation and 1200 bps VHF packet.
- PK-12 offers 1200 bps VHF packet at less than 80 mA of power.
- PK-96 features true DCD state machine for open squelch operation.
- 32K RAM, easily expandable to 128K.
- KISS mode for TCP/IP compatibility.
- Control of third-party MailDrop traffic.

Call (800) 432-8873 for a complete catalog or information on a specific product. Contact your favorite amateur radio equipment dealer for best pricing.

æA

Connect with us

Advanced Electronic Applications, Inc.
2006 196th Street Southwest
P. O. Box C2160
Lynnwood, WA 98036
(206) 774-5554

VIDEO LIBRARY

TECH NOTES

Test gear for the ham shack/ and a modification

Hams who homebrew also like to test their creations. Here are two simple projects to add to your collection of test equipment.

Peter Bertini, K1ZJH
Senior Technical Editor

This simple resistive tap lets you connect transmitters directly to your spectrum analyzer for highly accurate spectral purity measurements.

A −30 dB, 0 to 500 MHz Signal Sampler for Your Spectrum Analyzer

Rick Littlefield, K1BQT

Spectrum analyzers have fallen in price to a point where many experimenters now own them. This handy tool provides a unique view of the RF environment—displaying the frequency and amplitude of primary signals, plus a host of unwanted spurs, harmonics, parasitics, and other events. For casual bench work, RF sampling by means of a pick-up antenna or sniffer probe is usually sufficient to "see" what's going on. However, these proximity samples (samples taken with no direct connection to the signal source) are rarely accurate. Serious assessment of transmitter performance requires a better method of coupling signals into the spectrum analyzer.

Direct Sampling

To avoid ambiguity, the signal path between transmitter and analyzer should be direct. This way, sample errors caused by conditions such
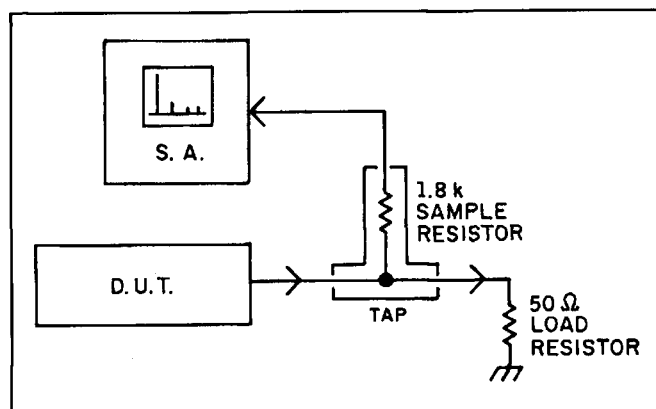
Figure 1. Diagram illustrates direct signal path from transmitter output to spectrum analyzer through use of a resistive tap.

as multipath, selective pick-up response, mismatch, and RF leakage will be minimized or eliminated. Here are some special considerations for direct hookups:

Shielding: Even small amounts of RF leakage may introduce spurious pickup that can distort results. Double-shielded coax, high-quality connectors, and RF-tight attenuator enclosures provide essential safeguards against RF leakage.

Attenuation: Most analyzer front-ends are designed to measure signals in the milliwatt range. This makes it necessary to attenuate the transmitted signal down to a level the analyzer can measure.

Flat Frequency Response: In order to get an accurate picture of transmitter performance on the CRT display, the sampling network must exhibit flat frequency response over the entire range of interest. Failing this, you must provide an attenuation curve to correct the displayed CRT values. Needless to say, flatter is better!

Attenuator Design

Although there are many high-accuracy RF-power attenuators on the commercial RF market that satisfy the conditions outlined above, they tend to be expensive. This prompted me to conjure my own. RF power attenuator circuits are generally very simple, but their mechanical design and construction require care. In the world of RF above 50 MHz, a T-pad stuffed into a mini-box won't do the job!

For the sake of expediency, I decided to build a resistive line-tap type attenuator (see **Figure 1**). This approach would allow me to save on construction time by using an inexpensive "store-bought" dummy load to terminate the transmitter and dissipate its output power. The Radio Shack RS-21-549 dummy load is especially well suited for this purpose. Small in size, it features 100 percent shielding, handles up to 100 watts, and exhibits low VSWR through 500 MHz. The same load has been marketed under other brand names and may be available from other distributors.

The line tap itself uses a common 1/4-watt carbon-film resistor to extract a sample signal from the terminated transmitter line. This resistor, which is relatively high in value compared to the 50-ohm line load, serves two functions. First, it attenuates the RF sample to a usable level. Second, it isolates the sample line from the load line (ideally, you should be able to open or short the sample line without affecting VSWR at the transmitter). After running some preliminary checks, I settled on a resistor value

of 1.8 k. In practice, this yielded around 30 dB of attenuation and provided excellent port isolation. To ensure 100 percent shielding and low VSWR, I installed the sampling resistor inside a short length of brass hardline—replacing the inner conductor with the resistor body.

Construction

Fabrication of the tap assembly takes a bit of "RF plumbing," but common hobby tools and a stout soldering iron are all you'll need to do the job. As illustrated in **Figure 2**, the sampling tap is built around a modified BNC T-Adapter (Radio Shack RS-278-111). In addition to this, you'll need a BNC-type UG-306 female-to-female coupler (RS-278-115) and a length of 13/32-inch thin-wall brass tubing (K&S no. 136, available at most hobby stores). If you use a "T" or coupler made by another manufacturer, you may need a slightly larger diameter tube to form the attenuator-line segment.

I used the following steps to assemble my taps. Please refer to Figure 2 throughout this operation for visual reference:

1. Locate the BNC "T" connector. With a fine-tooth hobby saw (or band saw), remove the top BNC connector barrel—cutting about 1 mm above the block to leave a small sleeve. Brass tubing will be slipped over this sleeve and soldered in place at a later time. Discard the cut-off barrel (it's too short to reuse) and deburr the opening in the block.

2. With a hobby saw or small tube cutter, cut the female-to-female coupler in half. Remove the inner conductor and insulator assembly from one half and deburr the barrel with an X-acto™ knife. Also, deburr the outside with a file or hand-grinder. Avoid removing plating from the barrel. You may set aside the other coupler half for later use on a second tap.

3. Cut a 5/8-inch length of 13/32-inch brass tubing (cut must be even all the way around). Deburr with an X-acto™ knife or hand-grinder, then clean both ends with fine emery paper inside and out. When "dressed," this tube should slip over the 1-mm flange on the "T" block and seat squarely. It should also accept the coupler barrel at the opposite end (check this before attempting final assembly).

4. Find a 1/4-watt 1.8-k carbon film resistor. Cut each lead to within 2 mm of the resistor body. Make a 180-degree crimp-bend at 1 mm in each lead and carefully tin with solder.

5. Locate the inner conductor and insulator assembly you removed from the coupler barrel. Extract the metal inner conductor pin from its insulator. Holding the pin with hemostats, melt a dab of solder on the inner's cut-off end and tack-solder the sampling resistor in place.

6. In similar fashion, tack-solder the free end
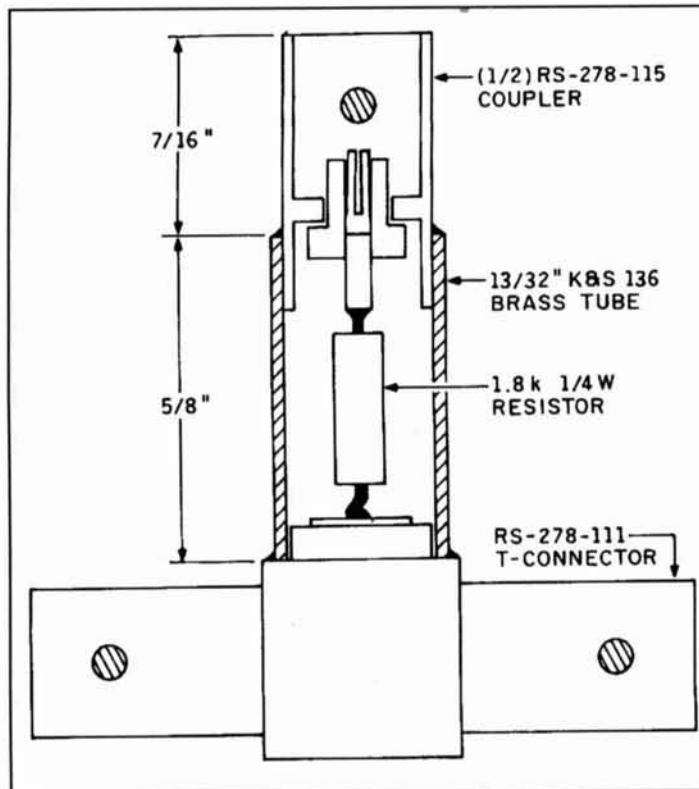
Figure 2. X-ray view of BNC "T" shows substitution of inner conductor with 1.8-k sampling resistor. Refer to this drawing during tap assembly.

of the resistor to the exposed inner conductor surface on top of the "T" block. Now, reinstall the plastic BNC insulator on the opposite end of the resistor. Your "T" assembly should look like the sample shown in **Photo A**.

7. Identify the cut-off end of the coupler barrel. Friction-fit this into one end of the prepared brass tube. Gently fit the other end of the brass tube down over the sampling resistor and seat it onto the "T" block. At the top, the coupler bar-
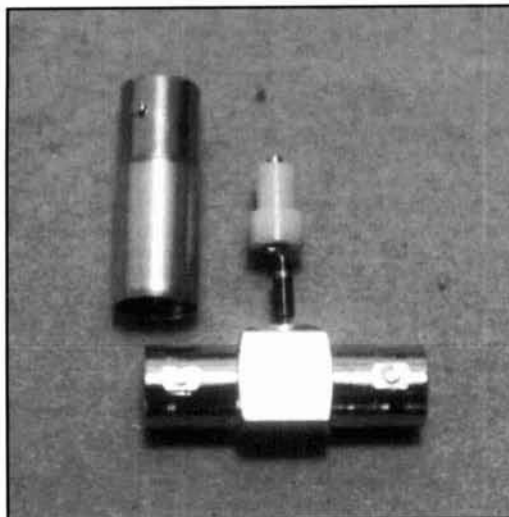
Photo A. Tap assembly prior to installation of outer conductor and connector assembly.
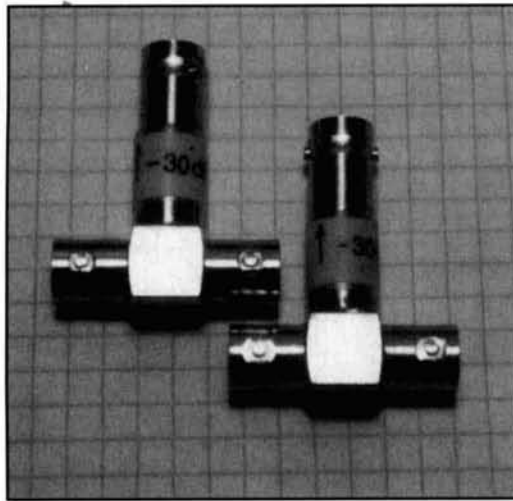
Photo B. Completed taps. Attenuation level and location of sample port should be noted on the tap to prevent confusion or misuse at a future time.

rel should protrude at least 7/16 inch above the brass tube. This will ensure proper seating of BNC connectors when the tap is completed.

8. Support the "T" block for soldering. With a high-wattage iron or small butane torch, solder the brass tube to the "T" block using standard rosin-core radio solder. Work quickly and be careful not to overheat the block. The specified Radio Shack connectors do not use Teflon™ insulation, which means the insulation may melt or distort if exposed to soldering temperatures for too long a period. Heat the block, flow solder around the base of the tube, then use forced air (or clamp in a metal vise) to cool the block as quickly as possible.

9. Following the same overheating precaution, sweat-solder the coupler barrel in place at the opposite end of the brass tube.

10. Clean all metal surfaces with acetone or other flux remover. File off excess solder, then
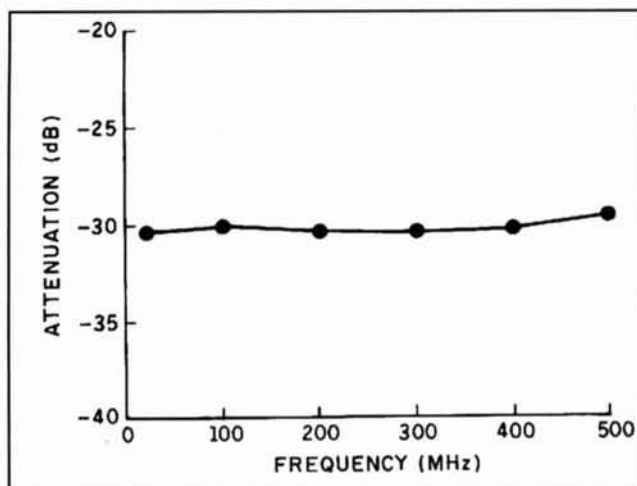
Figure 3. Frequency response of resistive tap and dummy load at the 0-dBm input power level. Measurements were made using a HP-608E signal generator and Ailtech 757 spectrum analyzer.

sand and buff as needed to finish (a hand grinder is helpful for this task). **Photo B** shows the finished product.

11. Using an ohmmeter, measure resistance between the load port and sample port to confirm that your 1.8-k sampling resistor survived assembly (usually not a problem).

Maximum Power Rating

With a 1.8-k sampling resistor installed, your tap should provide around 30 dB of RF attenuation (see **Figure 3**). This means a 100-watt input signal will produce a 100-mW sample at the tap (+50 dBm -30dB = +20dBm). Note that 100 watts is probably the absolute upper power limit for both the tap and the dummy load. While the sampling resistor is rated for 250 mW at DC, I would assume a significantly lower rating at VHF. Also, the prescribed dummy load will only tolerate 100 watts for a very short period of time. Be sure to read the load's derating instructions and carefully limit transmission time to prevent overheating. Finally, confirm that your analyzer front-end can handle 100 mW before attempting to measure signals in this power range. Transmitters in the 10 to 15 watt range may be tested safely without danger of overheating the load or burning out the sampling resistor.

Frequency Response

As mentioned earlier, flat frequency response over the sample range is essential for painting an accurate picture on your analyzer's CRT display. As **Figure 3** illustrates, the tap and dummy load described in this article test virtually flat out to 500 MHz. It is important to remember that the performance of your tap is totally dependent upon the flatness of your dummy load. Any load exhibiting significant VSWR in the frequency range of the sample will probably introduce error. For best results, I recommend keeping the sample line between the tap and your analyzer short. I generally connect mine directly to the analyzer's input jack using RF adapters.

Sample-Line Termination

Note that the sample port of the tap is not internally terminated. This is because most analyzers are already self-terminated at their input jack (50 ohms for lab analyzers and 75 ohms for cable-TV analyzers). If your analyzer isn't terminated for some reason, install a 50-ohm –3 dB inline pad between the tap and the analyzer input. Any time the sample line is unterminated (or inadvertently double-terminated), the accuracy of your readings may suffer because of degraded frequency response.

Conclusion

This simple device was built purely out of necessity; I claim no expertise as an RF-instrumentation engineer. Nevertheless, the tap seems to work well—and it has done much to improve the accuracy of purity measurements I make with my analyzer. It's a good feeling to be able to evaluate the performance of filters and other transmitter circuitry and fully trust the results!

As a side note, now that more experimenters are using spectrum analyzers, I eagerly look forward to reading a new wave of exciting articles describing ways to use them better (this is an unveiled hint to prospective *Communications Quarterly* authors).

Here's a solid-state version of the old "valve"-type grid dip oscillator that was a favorite of a ZL-Land ham.

FET Dip Oscillator

John R.L. Walker, ZL3IB
Reprinted from Break-In
January/February 1993

If one goes back to the valve era, the grid dip oscillator or GDO was one of the most useful items of test gear in the ham shack. For a long time, I have been interested in the design and construction of a transistorized dip oscillator that could perform as well in all respects as did the old valve GDO. In my experience, the solid-state units lacked frequency range, lacked output and, worst of all, exhibited many spurious dips as one tuned across the band.

However, I have found the following dual FET circuit to be free from the above defects and to be as good as my old valve GDO. Nothing original is claimed for the basic design, which appeared originally as a VHF dip oscillator in several RSGB publications.[1,2] An HF version was published by VK2ZCW[3] and my design is based on these circuits.

The oscillator uses a pair of FETs in a solid-state version of the "Kalitron" circuit which used a pair triodes as a VHF oscillator. Non-symmetrical loading of the oscillator is avoided by the use of a balanced diode detector which then feeds a simple DC amplifier to allow the use of a less sensitive meter. A phone jack is included in the metering circuit to allow the unit to be used as a simple monitor/wavemeter (in the DIODE mode). This function is controlled by the switch at point "C" (see **Figures 1** and **2**). In addition, a low level AM modulation facility is provided and this is switched by

Elliott D. Kaplan
The MITRE Corporation
Bedford, Massachusetts

THE GLOBAL POSITIONING SYSTEM (GPS)

An overview

The opinions of this paper are those of the author and do not reflect the views of The MITRE Corporation.

Functional description

GPS is a space-based radio navigation system that provides worldwide three dimensional position, velocity, and Universal Coordinated Time (UTC) information to users equipped with GPS receivers. Developed by the Department of Defense (DOD), the system was primarily designed for military use; however, the number of civilian users is expected to exceed those in the military. The system provides two services: the Standard Positioning Service (SPS) for the civilian community and the Precise Positioning Service (PPS) for the use by the military and select government agencies (e.g., the Central Intelligence Agency, Drug Enforcement Agency, and Federal Bureau of Investigation). PPS access is controlled through cryptography. GPS implementation and operational policies are delineated in the Federal Radionavigation Plan (FRP), prepared jointly by both the DOD and the Department of Transportation (DOT).

Principle of operation

Let's begin our discussion of GPS with a look at the principle of operation. In **Figure 1**, we wish to determine vector $\underline{U}$, which represents a user receiver's position with respect to

the center of the Earth. Vector $\underline{R}_S$ represents the distance and angular orientation from the center of the Earth to the satellite. Vector $\underline{R}_S$ is calculated using ephemeris (position) data broadcast by the satellite. The user-to-satellite distance vector, $\underline{\rho}$, is:

$$\underline{\rho} = \underline{R}_S - \underline{U} \tag{1}$$

Thus, the magnitude of vector $\underline{\rho}$, is:

$$\|\underline{\rho}\| = \|\underline{R}_S - \underline{U}\| \tag{2}$$

Let, ρ represent the magnitude of vector $\underline{\rho}$:
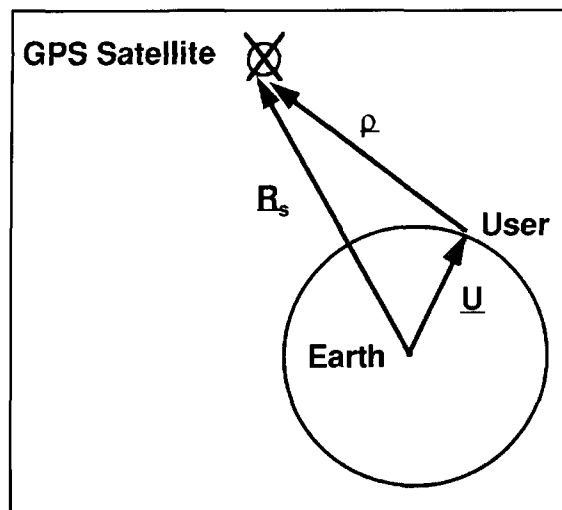
Figure 1. User position vector representation.
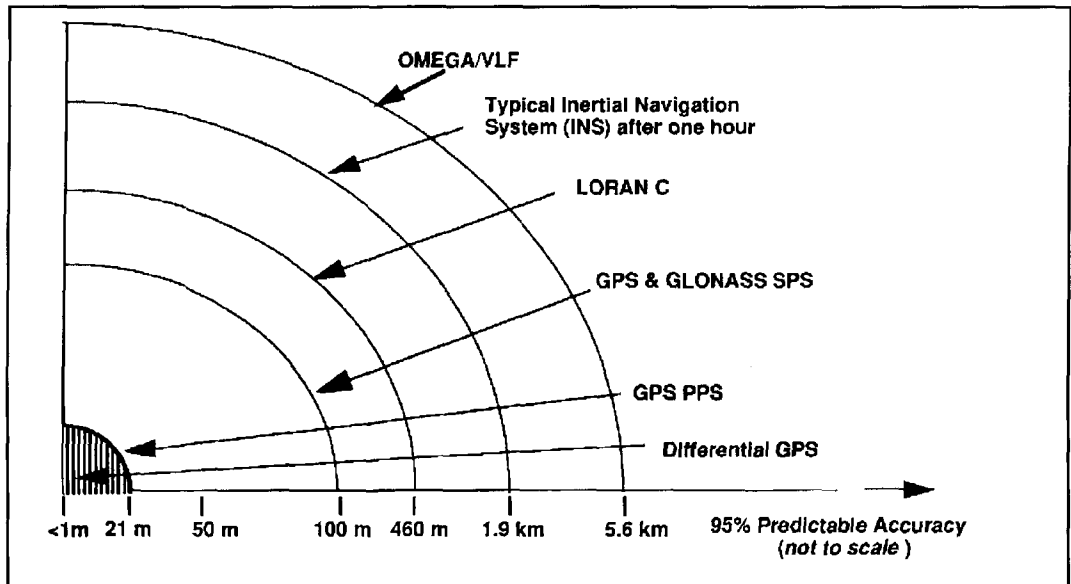
Figure 2. Accuracy comparison of GPS with other navigation systems.

$$\rho = \|\underline{R}_s - \underline{U}\| \tag{3}$$

Distance ρ is computed by measuring the propagation time required for a satellite-transmitted ranging code to reach the receiver. Within the receiver, an identical coded ranging signal is generated and shifted in time until it achieves correlation with the satellite-generated ranging code. The magnitude of the time shift is the satellite-to-user range. If the satellite clock and the receiver clock were perfectly synchronized, the correlation process would yield the true satellite-to-user range; however, this is not the case. Each satellite contains a highly accurate cesium or rubidium atomic clock assumed to be synchronized to GPS system time.* (GPS system time is an internal reference that is transparent to the user.) The receiver contains a less accurate crystal clock set approximately to GPS system time. Thus, the range determined by the correlation process is denoted as the pseudorange, R, because it contains the true satellite-to-user range, ρ, and an offset attributed to the difference between GPS system time and user clock. Therefore, **Equation 3** can be rewritten as:

$$R - c\Delta t = \|\underline{R}_s - \underline{U}\| \tag{4}$$

where c is the speed of light and Δt represents the advance of the receiver clock with respect to GPS system time.

To determine user position in three dimensions (U_x, U_y, and U_z) and the offset Δt, pseudorange measurements are made to four satellites resulting in:

$$R_i - c\Delta t = \|\underline{R}_{si} - \underline{U}\| \tag{5}$$

where i denotes satellites 1, 2, 3, and 4.

This equation can be expanded into the following set of equations, with four equations and four unknowns (U_x, U_y, U_z, and Δt):

$$R_1 - c\Delta t = \sqrt{(X_1 - U_x)^2 + (Y_1 - U_y)^2 + (Z_1 - U_z)^2} \tag{6}$$

$$R_2 - c\Delta t = \sqrt{(X_2 - U_x)^2 + (Y_2 - U_y)^2 + (Z_2 - U_z)^2} \tag{7}$$

$$R_3 - c\Delta t = \sqrt{(X_3 - U_x)^2 + (Y_3 - U_y)^2 + (Z_3 - U_z)^2} \tag{8}$$

$$R_4 - c\Delta t = \sqrt{(X_4 - U_x)^2 + (Y_4 - U_y)^2 + (Z_4 - U_z)^2} \tag{9}$$

where X_i, Y_i, and Z_i denote the ith satellite's position in three dimensions.

Several closed form solutions published in the open literature or various iterative techniques such as Newton-Raphson are used to solve these nonlinear equations for the unknown quantities.[1]

Comparison of GPS to existing navigation systems

The benefits of GPS are readily observed

*Satellite clocks are free running and each has a unique time bias from GPS system time. However, these time biases are computed by the GPS ground monitoring network (described later) and are uplinked to the satellites for rebroadcast to the users. Because the corrections are applied within the user receiver to synchronize each ranging signal to GPS time, the assumption is considered valid.

1575.42 MHz
L1 Carrier

$\cos(\omega_1 + \theta)$

$\Delta\phi = 90°$ $\sin(\omega_1 + \theta)$

1.023 Mbps
C/A-Code $C(t)$

50 bps
Message $D(t)$

$D(t)\,C(t)$

$D(t)\,P(t)$

10.23 Mbps
P-Code $P(t)$

$S_{L1}(t) = D(t)\,C(t)\sin(\omega_1 + \theta)$
$+$
$D(t)\,P(t)\cos(\omega_1 + \theta)$

1227.6 MHz
L2 Carrier $\cos(\omega_2 + \theta)$

$S_{L2}(t) = D(t)\,P(t)\cos(\omega_2 + \theta)$

10.23 MHz
Oscillator

⊕ **Modulo-2 Adder**   ⊗ **BPSK Modulator**   Σ **Signal Combiner**
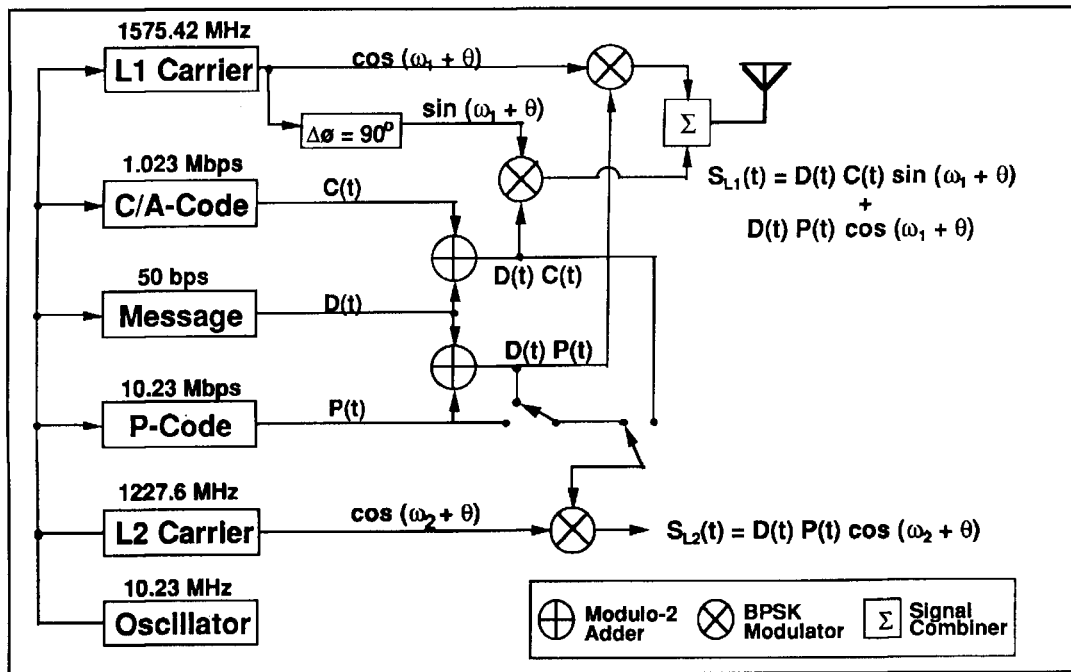
Figure 3. Satellite signal generation.

when compared to some of today's widely used navigation systems. Key features of these systems—such as accuracy, coverage, and frequency—are provided below. An accuracy comparison of these systems is provided in **Figure 2**.

OMEGA. Aside from GPS, OMEGA is the only radio navigation aid providing worldwide coverage. (The TRANSIT satellite system provides two-dimensional worldwide coverage, but DOD plans to terminate TRANSIT operations in 1996.) OMEGA is jointly operated by the DOD and six other nations. Broadcasting very low frequency (VLF) CW signals in the frequency region of approximately 10 kHz to 14 kHz, OMEGA stations provide horizontal guidance only. Predictable horizontal position accuracy varies from 3.7 km to 7.4 km, (95 percent). According to the FRP, the DOD is expected to discontinue OMEGA usage in 1996. OMEGA civilian operations are expected to continue until 2005.[2]

LORAN-C. LORAN-C is a radio navigation system widely used for civilian/military aviation, land, and marine applications. Unlike OMEGA, LORAN-C coverage is regional (e.g., contiguous United States, lower sections of Canada, portions of Japan, etc.). LORAN-C transmissions are at 100 kHz and provide horizontal guidance with a predictable accuracy of 460 m. DOD is discontinuing its LORAN-C requirement starting in 1995. Civilian use of LORAN-C within CONUS is expected to continue until 2015. The U.S. Coast Guard will terminate overseas LORAN-C operations at the end of 1994.[2]

Inertial Navigation. Unlike LORAN-C and OMEGA, an inertial navigation system (INS) is not a radionavigation system, but rather an electromechanical mechanism that provides navigation guidance based on Newton's Laws of Motion. An INS uses gyroscopes, accelerometers, and associated electronics (including a computer) to sense turning rates and accelerations associated with the rotation of the platform. The accelerometers measure individual horizontal and vertical acceleration components, while the gyroscopes stabilize the accelerometers in a desired orientation. These accelerations are integrated by the computer to obtain platform position and velocity. The computer also calculates orientation corrections caused by the motion over the Earth, the rotation of the Earth, and other factors. INS position and velocity accuracies degrade over time. A nominal position drift error rate is 1.9 km/hour (95 percent). If the INS is installed in an integrated navigation system with one or more external navigation aids, such as GPS and/or LORAN, the external navigation aid can be used to update the INS and correct the drift error.[3]

Global Navigation Satellite System (GLONASS). GLONASS is the Russian counterpart to GPS. GLONASS provides worldwide coverage, however, its accuracy performance is optimized for the northern latitudes (Western Europe and Russia, for example). GLONASS accuracy performance is specified as identical to that of GPS SPS.

GPS SPS. As stated earlier, GPS SPS is the primary service provided to the civilian com-
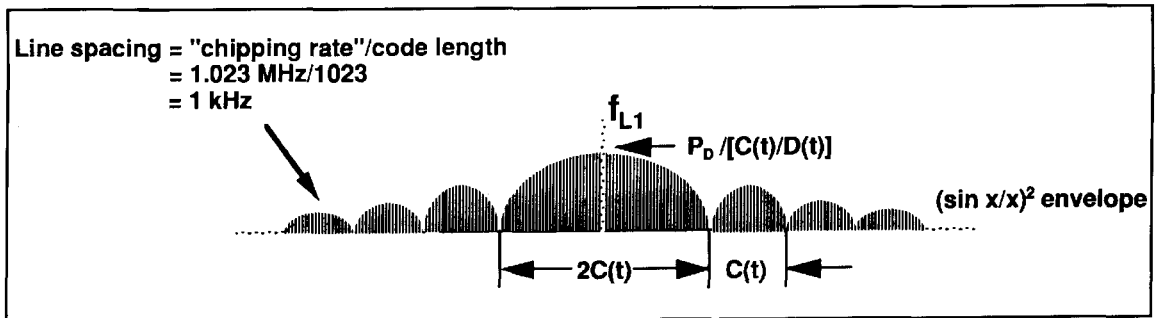
Figure 4. Effects of spreading data.

munity worldwide. The service provides accuracies on a daily basis within 100 m (2 drms,* 95 percent probability) in the horizontal plane and 140 m (95 percent probability) in the vertical plane. Further, the UTC output is within 340 nsec (95 percent probability).[2]

GPS PPS. The GPS PPS is designated for authorized users (for example, select government agencies and the military) equipped with cryptographic capabilities. On a daily basis, PPS horizontal accuracy is 21 m (2 drms, 95 percent probability) in the horizontal plane and 29 m (95 percent probability) in the vertical plane. The UTC output is within 200 nsec (95 percent probability).[2]

Differential GPS (DGPS). DGPS is an accuracy enhancement technique that provides horizontal and vertical positioning accuracies better thanthose of the PPS. Further elaboration on DGPS is provided later in the text, but it is important to note that some DGPS techniques provide measurement accuracy on the order of millimeters.

GPS system components

GPS is comprised of three segments: space, operational control, and user. A brief description of each segment is provided below:

Space Segment. The space segment consists of 24 satellites positioned in six orbital planes with each plane containing four satellites. Satellite orbit altitude is about 20,186 km (10,900 nmi) above the Earth's surface. At present, the constellation consists of 24 Block (BLK) II/IIA (production) satellites and 1 BLK I (prototype) satellite. The system reached initial operation capability (IOC) in mid-December 1993. IOC marked the availability of

the SPS. The full operational capability (FOC) constellation is in place now and consists of the 24 BLK II/IIA satellites. The last BLK II/IIA production satellite was launched in March 1994. Declaration of FOC will follow an extensive system "check-out" and is expected sometime in early 1995. Once FOC is attained, four to nine satellites will be in view at any time. (The number of satellites in view is a function of user location, time of day, and satellite outages.) BLK IIR replenishment satellites are in production. The first launch is anticipated for 1996. This launch date is based on BLK II/IIA reliability predictions. A BLK IIF follow-on program for the replacement of the BLK IIR series is being planned.

Operational Control Segment (OCS). The OCS consists of a master control station (MCS) at Falcon AFB, Colorado and monitor stations in the following locations: Ascension, Diego Garcia, Falcon AFB, Hawaii, and Kwajalein. A backup MCS is located at Gaithersburg, Maryland. The OCS monitors the health and status of the constellation and uploads navigation data to the satellites for retransmission to users. The monitoring stations passively track and collect data from all GPS satellites in view. This data is relayed every 6 seconds to the MCS via the Defense Satellite Communications System (DSCS). MCS data processing examines monitoring station data for anomalies and computes new navigation parameters (for example, almanac [i.e., coarse satellite position data], ephemeris [i.e., fine satellite position data], health, satellite clock corrections, etc.). The new navigation parameters are disseminated to all monitoring stations via DSCS and uploaded to the constellation via an S-band link. Only three monitoring station sites (Ascension, Diego Garcia, and Kwajalein) have an upload capability.

User segment. The user segment consists of GPS receiving equipment. Today GPS receivers come in a variety of forms ranging from a single board of "credit card" size to shock-mounted Navy shipboard units with a "footprint" of approximately 1500 sq. cm that weigh 31.8 kg (70 lbs.) each. Single-board receivers

---

*2 drms is the radius of a circle centered about the receiver equal to twice the root-sum-square of the variances around the x and y axes:

$$2 \text{ drms} = \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (10)$$

The percentage of fixes contained within the 2 drms circle varies between approximately 95.5 percent and 98.2 percent, depending on the ellipticity of the error distribution. For GPS, the 95 percent value is used.[2]
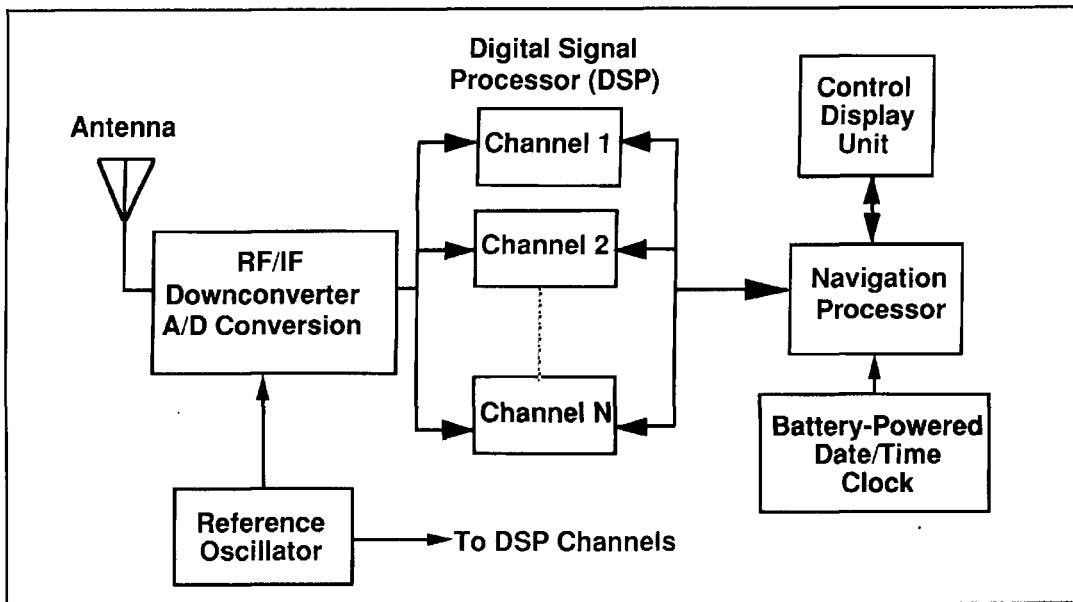
Figure 5. Generic SPS GPS receiver.

are available from several manufacturers. These devices are primarily for SPS OEM applications and require an antenna, display, I/O electronics, and power. There are also numerous self-contained SPS handheld receivers on the market designed for aviation, land, and marine applications. Aviation receivers are designed for a high dynamic environment. Rockwell-Collins is currently producing the PPS Precision Lightweight GPS Receiver (PLGR) for the DOD. Aside from handhelds panel mount receivers for aviation and marine applications are also popular. Unlike handhelds, panel mount receivers rely on platform power rather than batteries. In addition, panel mount units offer hands free operation and provide security in terms of a fixed mounting.

The majority of military airborne GPS receiver configurations are composed of several components: antenna, RF/IF downconverter assembly, receiver, and control/display unit. In some tactical platform configurations, an adaptive array antenna and associated downconverter provide an anti-jam capability. This array is electronically steerable and creates nulls in the antenna pattern in the direction of the jammer.

GPS satellite signal characteristics

Satellite signal generation is illustrated in **Figure 3**. Each satellite contains a 10.23-MHz atomic clock (oscillator) that provides reference timing for all onboard signals. (Depending on the spacecraft type, there can be up to four reference oscillators on board; however, only one
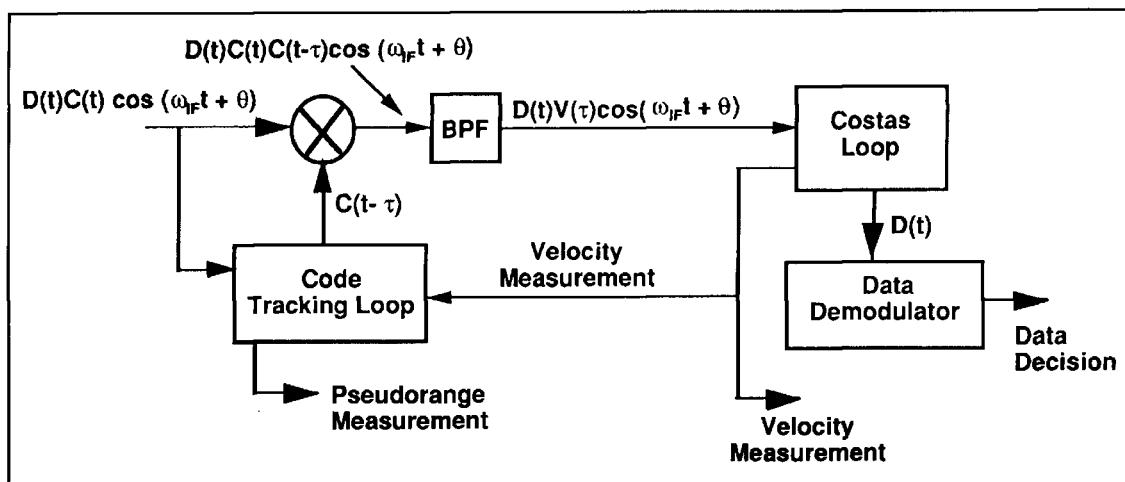
Figure 6. Direct sequence "despreading" technique.

Amateur Radio and GPS

A compact Global Positioning System (GPS) unit can be a handy piece of equipment for the amateur radio operator who likes to use the grid locator system popular with VHF+ aficionados. Introduced in Germany over 40 years ago to encourage activity on the VHF+ ham bands, the grid locator scheme evolved into a worldwide awards and location system. Today's program is called the Maidenhead Grid Locator System, and is used daily by hams all over the world to describe their location.

Grid locator designators are determined by longitude and latitude. The designator itself contains up to six places—two letters followed by two numbers, followed by two letters. The grid system divides the Earth's surface into 324 "fields" of 20 degrees of longitude and 10 degrees of latitude. The fields are identified by two letters—AA through RR. Each field is then divided into 100 "squares" of 2 degrees of longitude x 1 degree of latitude; these are identified by two numbers from 00 to 99. Each square is further divided into 576 subsquares of 5 minutes of longitude x 2.5 minutes of latitude. Subsquares are identified by two letters (AA-XX). The trick for the amateur radio operator is to figure out, as accurately as possible, his location within the grid system.

While it's certainly possible to determine your particular grid locator designator using tools like the *ARRL Grid Square Map for North America* and the *ARRL World Grid Square Atlas*; shareware software like *BD, HAMGRID, LOCCALC,* and *SQUARES*; or U.S. Geological Survey maps; these inexpensive methods are not always the easiest. Enter the handheld GPS receiver. This little gadget, though considerably more expensive, lets you obtain the information you need to find your grid locator designator quickly and with a fair degree of accuracy. GPS handheld units rely on the same technology and 24-satellite system as their larger, more complex counterparts. A unit's accuracy is based on the number of satellite signals it receives. The GPS receiver must receive a minimum of three satellites to obtain the signal necessary to produce a two dimensional location, which will allow the circuitry in the unit to calculate your location. However, it takes a minimum of four satellitesto determine accurate elevation .

Several manufacturers offer handheld GPS units. Prices range from $500 and up. One such unit is shown in **Photo A**. For more information on handheld global positioning systems check out the catalogs of amateur radio equipment suppliers and manufacturers of navigational aids.

Terry Littlefield, KA1STC
Editor

BIBLIOGRAPHY
1. Joseph Lynch, N6CL, *The VHF "How To" Book*, Chapter 15, CQ Communications, Inc., 1994.
2. Arthur Lange, W6RXG, David Sprague, and Arthur Woo, "Scout GPS ™ Application Note 6," Trimble Navigation, Austin Texas. To obtain a copy, call 1-800-959-9567

Photo A. The Trimble Scout GPS ™ handheld unit has the Maidenhead Grid Locator System built into its software. The unit's RS-232I/O port lets you upload and download location data to and from your computer.

SEGMENT SOURCE	ERROR SOURCE	P-CODE PSEUDO RANGE ERROR 1σ [m]
SPACE	Clock and nav, subsystem stability	3.3
	Other	0.5
CONTROL	Ephemeris prediction model implementation	4.2
	Predictability of satellite perturbations	1.0
	Other	0.9
USER	Ionospheric delay	2.3
	Tropospheric delay compensation	2.0
	Receiver noise and resolution	1.5
	Multipath	1.2
	Other	0.5
SYSTEM UERE	Total (Root-sum-square)	6.6

Figure 7. P(Y) code UERE budget.

SEGMENT SOURCE	ERROR SOURCE	C/A-CODE PSEUDO RANGE ERROR 1σ [m]
SPACE	Selective Availability	30.3
	Clock and nav, subsystem stability	3.3
	Other	0.5
CONTROL	Ephemeris prediction model implementation	4.2
	Predictability of satellite perturbations	1.0
	Other	0.9
USER	Ionospheric delay	5.0
	Tropospheric delay compensation	0.8
	Receiver noise and resolution	1.0
	Multipath	1.2
	Other	0.5
SYSTEM UERE	Total (Root-sum-square)	31.3

Figure 8. C/A code UERE budget (estimate).

is used as the primary timing source. The others provide redundancy.) The reference timing signal is upconverted (multiplied by 154) to create the L1 carrier at 1575.42 MHz. The L1 carrier is designated as the primary navigation frequency. The 10.23-MHz source is also multiplied by 120 to form the L2 carrier at 1227.6 MHz. The L2 carrier is designated as the secondary navigation frequency. Each satellite broadcasts on L1 and L2. Carrier L1 is separat-

ed into quadrature components. The in-phase and 90-degree phase-shifted components are represented by $\cos(\omega_1 t+\theta)$ and $\sin(\omega_1 t+\theta)$, respectively. Here ω_1 denotes the L1 carrier frequency and θ represents oscillator phase noise and drift. In addition to carrier frequency generation, the 10.23-MHz source is also used to form the coarse/acquisition (C/A) and precision (P) ranging code signals, and the navigation data message (NAV data message). Both

ranging codes appear to consist of a random pattern of binary digits and are denoted as pseudorandom codes. The pseudorandom code ones and zeros are equal in magnitude, but reversed in sign. A one is usually represented by +1 volt, while a zero is generally denoted by a -1 volt.

The C/A (ranging) code is a 1023 bit Gold code clocked at 1.023 MHz. Thus, the code has a 1 msec epoch and repeats constantly. Each satellite is assigned a unique Gold code. Gold codes were selected for GPS because a subset of codes within the family have low cross correlation properties with respect to one another. This is an important property because the satellite broadcast format is code division multiple access (CDMA) with each satellite transmitting its C/A-code on L1. Thus, there are multiple C/A-codes on L1 simultaneously. In order to make a pseudorange measurement to a particular satellite, the receiver must be able to distinguish one code from another. C/A-codes are primarily used by the civilian community for SPS measurements.

The P (ranging) code is a 267-day sequence with a bit rate of 10.23 MHz. Each satellite is assigned a unique one-week segment of the P-code that repeats every midnight Saturday/Sunday. Each P-code segment has low cross-correlation properties with respect to other P-code segments and Gold codes. Once FOC is declared, P-code transmissions will be encrypted to form the Y-code. The Y-code will be accessible to authorized users through cryptography. Y-code cross-correlation properties are assumed to be identical to those of the P-code. (Y-code properties are classified and I have no knowledge of them.)

The NAV data message contains information regarding the health and status of the satellites, as well as data to aid in the satellite-to-user measurement process. NAV data message parameters include almanac, ephemeris, satellite clock corrections, GPS system time, UTC offset from GPS system time, and C/A-code to P(Y)-code handover information. Authorized receivers usually acquire the satellite transmission using the C/A-code and then switch to P(Y)-code tracking via the use of handover information. The NAV data message is broadcast at 50 Hz. Messages are 12.5 minutes in length and cyclic.

The NAV data message is modulo-2 added to both the C/A-code and the P-code forming products D(t) C(t) and D(t) P(t), respectively (see **Figure 3**). The D(t) P(t) product binary phase shift key (BPSK) modulates the in-phase component of the L1 carrier while product D(t) C(t) BPSK modulates the phase shifted component. This modulation technique is called direct sequence spread spectrum. The modulo-2 addi-

tion effectively "spreads" the 50-Hz data, D(t), across approximately the same amount of spectrum required by the multiplying ranging code. The ranging codes are also called "chipping" codes. The bit rate of the chipping code is referred to as the "chipping" rate. In addition, signal power density is decreased by the ratio of the chipping rate to the data rate. **Figure 4** depicts the spectrum of the C/A-code chipping the NAV message data on L1. The carrier has been suppressed due to the double balanced mixer and balanced spreading code. That is, the C/A-code has an approximately equal number of ones and zeros. This property also resides in the P-code. The outputs of both BPSK modulators are summed and fed to a right-handed circularly polarized (RHCP) antenna for broadcast-to-user receivers. This signal transmitted by the ith satellite on L1 is:

$$S_{L1,i}(t) = P_i(t)D(t)\cos(\omega_1 t + \theta) + C_i(t)D(t)\sin(\omega_1 t + \theta) \qquad (11)$$

where:

$P_i(t)$ = P code on ith satellite
$C_i(t)$ = C/A code on ith satellite
$D_i(t)$ = NAV data message on ith satellite

L2 carrier modulation is similar; however, the modulation format is selectable by the OCS. Either the D(t) C(t) product or P(t) with or without the modulo-2 addition of D(t) is chosen to modulate the L2 carrier. For the most part, the D(t) P(t) product is employed. The modulated carrier is broadcast to user receivers through a different RHCP antenna than the one used for the L1 signal broadcast. The L2 signal transmitted from the ith satellite is:

$$S_{L2,i}(t) = P_i(t)D(t)\cos(\omega_2 t + \theta) \qquad (12)$$

Receiver measurement process

Figure 5 is a block diagram of a generic SPS GPS receiver. Satellite signals are received via the antenna, which is RHCP and provides near hemispherical coverage. Typical coverage is 160 degrees with near unity gain. In general, SPS receivers process only C/A-code on L1 because the P-code on both L1 and L2 will be encrypted once the system reaches FOC. The CDMA signals are then downconverted to an intermediate frequency (IF). Following a second downconversion to baseband, the signals are sampled by an analog-to-digital converter at a frequency greater than or equal to twice the C/A-code bandwidth (2.046 MHz). In some receivers, sampling occurs at the reference oscillator rate, 10.23 MHz. Because the samples aren't synchronized to the received C/A codes, oversampling ensures that the signal will be sampled

throughout and not mistakenly near a code chip transition.[4] The samples are forwarded to the digital signal processor (DSP). The DSP contains N parallel channels to track the carriers and codes from up to N satellites simultaneously. (N generally ranges from 5 to 12 in modern SPS receivers.) Each channel contains carrier and code tracking loops to perform pseudorange and velocity measurements, as well as NAV demodulation. Prior to satellite signal tracking, the navigation processor predicts which satellites should be available. This prediction is based on: a) stored almanac data, b) last known values of user position, velocity, and time (PVT), and c) receiver clock time. The receiver attempts to achieve code correlation with those satellites predicted by the almanac data. Since the almanac data predicts that a subset of the satellite constellation is available for navigation, only a subset of the internal ranging code generator's codes is used for satellite signal correlation.[4] If almanac data is not available and/or the initial PVT data is inaccurate, the receiver "searches the sky" by adjusting its code generator to acquire at least one satellite. Once this satellite is acquired, up-dated almanac parameters are obtained from the NAV message. The receiver uses this new data to predict which satellites should be in view.[5]

A block diagram of the carrier and code tracking loops for a single receiver channel is shown in **Figure 6**. The downconverted SPS signal, D(t) C(t) cos(ω_{IF}t +θ), is input to the mixer and code tracking loop. The code tracking loop generates a pseudorandom code, C(t-τ) identical to the one being received from the satellite. Time delay τ represents the timing offset between the satellite-generated code and receiver-generated replica code. Correlation is achieved when τ=0. The time shift required to
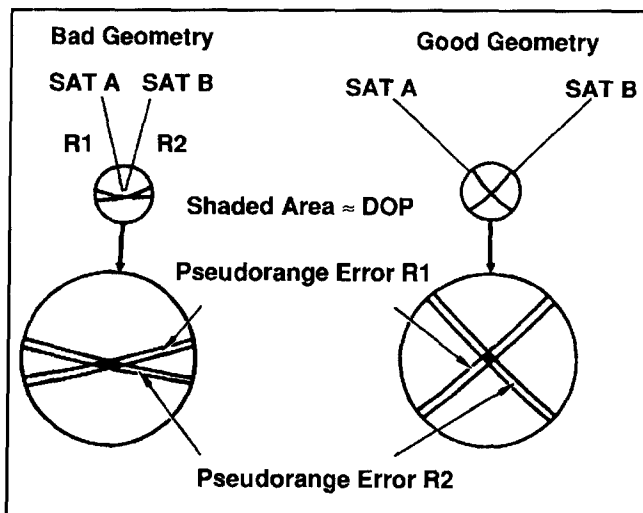
Figure 9. Dilution of precision.

align the code sequences is equal to the time required for the satellite signal to propagate to the receiver. This propagation time, multiplied by the speed of light, is the pseudorange. The pseudorange also contains measurement errors contributed by the clock bias error, initial code synchronization error, ionospheric and tropospheric delays, multipath, and other errors. (Elaboration on pseudorange errors appears in the following section.) After mixer multiplication, the product, D(t)C(t)C(t-τ) cos (ω_{IF} t + θ), is bandpass filtered. The bandpass filter is equivalent to an integrator and passes only the data modulation and residual doppler frequency uncertainty. The output of the bandpass filter is represented by D(t) V(τ) cos (ω_{IF} t +θ). Where V(τ) is the autocorrelation function of the code:

$$V(\tau) = Avg[C(t)C(t - \tau)] = 1 - \frac{\tau}{T} \qquad (13)$$
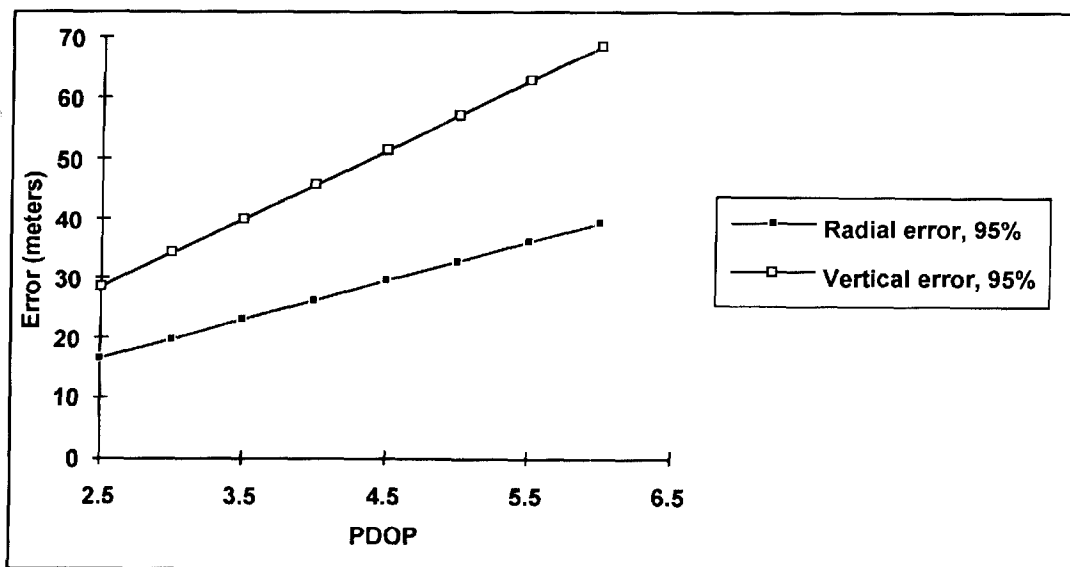
Figure 10. PPS horizontal and vertical error.
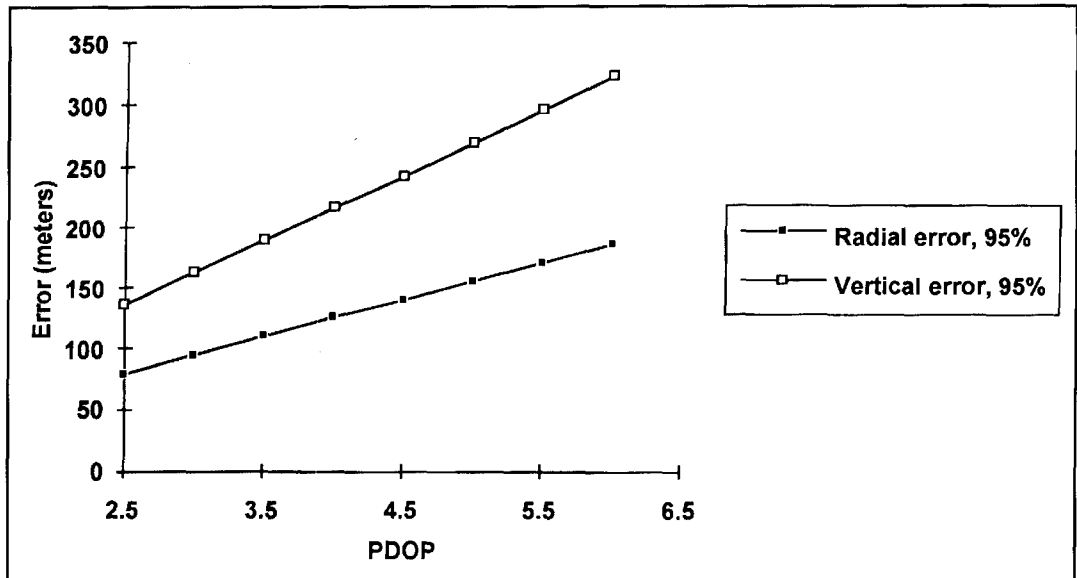
Figure 11. SPS horizontal and vertical error.

The output of the bandpass filter is applied to the Costas loop.

The Costas loop generates an intermediate frequency (IF) close to the expected IF. The difference between the downconverted carrier (expected IF) and the internally generated IF is the Doppler frequency associated with the relative velocity between the satellite and the receiver. It's important to note that this velocity includes the host vehicle dynamics. IF signal tracking occurs after code correlation because the despreading process recovers the carrier.

The Costas loop then adjusts the internally generated IF to match the downconverted carrier frequency.[5] This process yields the relative velocity between the satellite and the receiver. The receiver differences the adjusted IF with the downconverted carrier to obtain the magnitude of the Doppler frequency. This velocity data is used to aid the code tracking loop. NAV data is integrated over the 20 msec bit period within the Costas loop. The integrator output is forwarded to a data decision circuit where it's determined if the data is a one or a zero.

SEGMENT SOURCE	ERROR SOURCE	P-CODE PSEUDO RANGE ERROR 1σ [m]
SPACE	Clock and nav, subsystem stability	-
	Predictability of satellite perturbations	-
	Other	-
CONTROL	Ephemeris prediction model implementation	-
	Other	-
USER	Ionospheric delay	0.2
	Tropospheric delay compensation	0.2
	Alignment error	0.4
	Receiver noise and resolution	0.8
	Multipath	1.2
	Other	-
SYSTEM UERE	Total (Root-sum-square)	1.5

Figure 12. DGPS UERE budget (estimate).

Pseudorange and velocity measurements, as well as the NAV data from each channel, are input to the navigation processor for computation of the navigation solution (e.g., user position, velocity, and time).

The navigation processor contains a Kalman filter that computes optimal estimates of user position, velocity, and time. The filter is generally updated at a 1 Hz rate with new pseudorange and velocity measurements. Receivers designated for aircraft precision approach and other high dynamic applications have measurement Kalman filter update rates of at least 5 Hz. Filter navigation solutions are forwarded to a control display unit (CDU). The CDU permits operator data entry, display status and navigation solution parameters, and usually accesses numerous navigation functions such as waypoint entry, time-to-go, etc.

Performance Accuracy/Primary Error Sources

GPS measurement accuracy is a function of pseudorange measurement error and satellite geometry. Pseudorange measurement error is denoted as the user equivalent range error (UERE). UERE is modeled as a zero mean Gaussian random variable and is comprised of space, control, and user (receiver) segment error contributions. Descriptions of constituent error sources are provided below:

Selective Availability (SA). Under current DOD policy, SA is implemented to deny full system accuracy to unauthorized users (i.e., the civilian community). SA "dithers" each satellite's 10.23 MHz reference oscillator output, corrupting pseudorange measurement accuracy. SA also encrypts the NAV data message so key elements of the navigation data solution (e.g., satellite clock corrections and ephemeris) are reported inaccurately. Authorized users remove SA effects through cryptography.[5,6]

Satellite Clock Error. As mentioned earlier, the 10.23-MHz satellite reference oscillators are "free-running" cesium or rubidium atomic standards. Although these clocks are highly stable, they may deviate up to about 1 msec from GPS system time.[7] An offset of 1 msec translates to a 300 km pseudorange error. This effect is mitigated by the generation of satellite clock correction estimates by the OCS. Since these corrections are estimates of the actual satellite clock errors some residual error remains.

Relativistic Effects. Both Einstein's general and special theories of relativity are factors in the pseudorange measurement process. Relativistic effects result from differences in the gravitational potential between the satellite and user (general theory), as well as differences

in satellite-user relative velocity (special theory). The satellite is subjected to only six percent of the gravitation force felt at the Earth's surface; whereas, in most applications the user receiver experiences 100 percent of the Earth's gravitational force. In addition, the satellite velocity is approximately 3.7 km/sec (12,000 ft/sec) while the user receiver velocity is generally on the order of 0.3 km/sec (1,000 ft/sec), or less. Thus, the satellite clock "ticks" at a different rate than the receiver clock.[8] As a result, the satellite signal speeds up as it approaches the Earth causing the observed frequency to increase. To compensate for this time dilation effect, the satellite clock frequency is adjusted to 10.22999999545 MHz (10.23 MHz nominal) before launch.[9]

Ephemeris Prediction Model Implementation. Satellite ephemeris is determined by four OCS monitoring stations simultaneously making pseudorange measurements to a particular satellite. Each station is located at a surveyed position and contains an atomic clock referenced to GPS system time. This process can be considered an inverse GPS measurement system where the monitoring stations are the "satellites" and the satellite is the "user." Therefore, the same set of four simultaneous equations developed in the section on Principle of Operation can be used to solve for the satellite position.[7] At the MCS, optimal estimates of ephemeris data for all satellites are computed and uplinked to the satellites with other NAV data message parameters for rebroadcast to the user. As in the case of the satellite clock corrections, these corrections are estimates and contain a residual error.

Unmodeled Space Perturbations. Unmodeled space perturbations related to uncertainties in the solar radiation pressure and lunar-solar gravitational forces act on the spacecraft, contributing to ephemeris error.

Ionospheric delay. The time delay associated with the satellite signal wavefronts traversing the ionosphere is a function of the ionospheric total electron count (TEC), as well as ray bending. Both effects are due to refraction and are dependent on the character of the ionosphere and the satellite elevation angle. Almost three times as much delay is incurred when viewing satellites at low elevation than at the zenith. For a signal arriving at vertical incidence, the delay ranges from about 10 nsec at night to as much as 50 nsec during the day. At low satellite viewing angles (0 through 10 degrees), the delay can range from 30 nsec at night and up to 150 nsec during the day.[9] Severe delays can be incurred at the geomagnetic equator and the poles, or during magnetic storms caused by solar flares.

Ionospheric delays are inversely proportion-
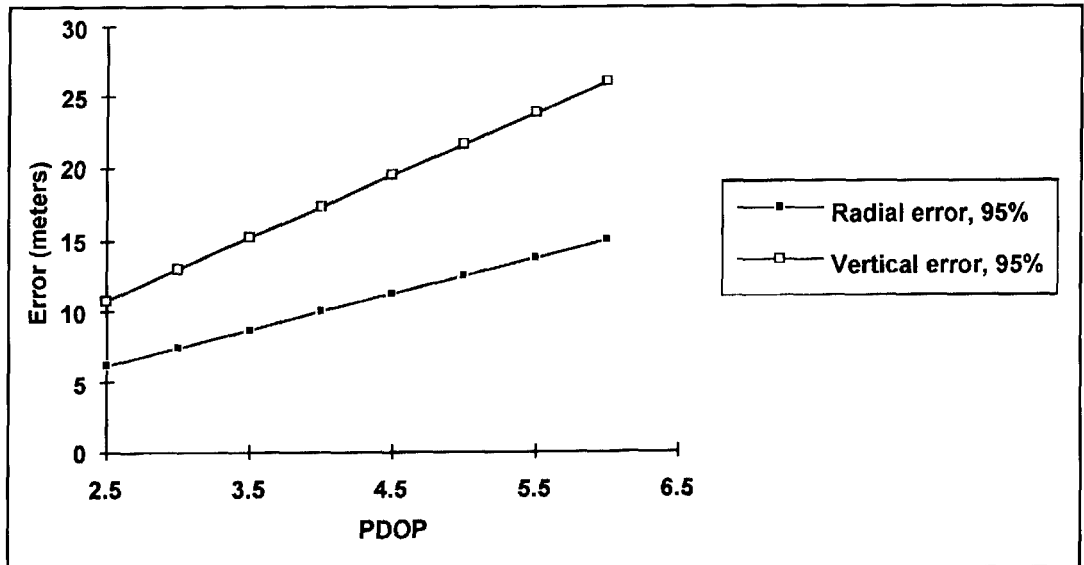
Figure 13. DGPS horizontal and vertical error.

al to the square of the transmitted frequency. As a result, the ionospheric delay can be estimated by making pseudorange measurements to the same satellite on both L1 and L2. The estimated difference in the corresponding time-of-arrival measurements, $\Delta\tau$, is used to calculate the ionospheric delay incurred by each frequency, τ_{dL1}, or τ_{dL2}.

$$\tau_{dL1} = \Delta\tau \ \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2} \tag{14}$$

$$\tau_{dL2} = \Delta\tau \ \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2} \tag{15}$$

These estimated delay corrections are subtracted from pseudorange measurements made by each frequency; however, because these corrections are only estimates, residual errors are usually present.

Most SPS receivers process C/A-code on L1 only. These receivers compensate for ionospheric delay with an onboard model that uses ionospheric delay correction broadcast in the NAV data message. **Reference 8** reports that the delay can be reduced by approximately 50 percent by employing the onboard compensation model.

Tropospheric Delay. Refraction from the satellite signal transiting the troposphere also induces a time delay into pseudorange measurements. The wavefront is refracted from water vapor and dry atmosphere. These effects are prevalent at low satellite elevation angles because the troposphere is somewhat denser closer to the Earth. Also, the delay is increased

when the user receiver is near the Earth's surface rather than at altitude. Compensation for tropospheric delay can be achieved with a mathematical model such as Essen and Froome[9] or Modified Hopfield.[10] Residual tropospheric delay errors contribute to pseudorange error.

Receiver Noise and Resolution. Receiver noise induces errors into pseudorange measurements. Most contemporary high-end receivers are designed to minimize measurement noise. Resolution refers to range error incurred within the ranging code correlation process.

Multipath. The GPS signal is reflected from various surfaces and structures surrounding the antenna. Depending on the angle of incidence and the surface of the reflecting material, reflected ray amplitudes can approximate that of the direct ray (i.e., desired signal). These reflected rays may be incident on the antenna and can combine with the direct ray. Within the receiver, reflected rays tend to distort direct ray ranging signals and induce measurement errors.

UERE Budget. The UERE budget for the P(Y)-code is shown in **Figure 7.** This budget was prepared by the GPS Joint Program Office (JPO) and serves as a guideline for all DOD GPS users.[5] (This budget was developed in the 1980s and has not been updated even though all aspects of today's space, control, and user segments have been improved.) Each budget component is modeled as a zero mean Gaussian random variable over a 24-hour period. Hence, the root-sum-squared (RSS) total of the Gaussian distributions remains Gaussian distributed with a 1-σ value of 6.6 m.

Figure 8 is an estimate of a contemporary SPS receiver C/A-code UERE budget. The

space and control segment contributions are identical to those in the P(Y)-code budget except for the inclusion of SA. SA is the dominant error source in SPS C/A-code pseudorange measurements. The value of SA used in this budget was derived from a UERE budget provided in **Reference 11**. With respect to the P(Y)-code budget, user segment error contributions differ for receiver noise and resolution, as well as atmospheric delay components. The value of ionospheric delay is an estimate that reflects delay compensation by an internal receiver model. I believe that today's SPS receiver manufacturers can obtain a better level of tropospheric delay compensation than reported in the 1980-era GPS JPO UERE budget. My estimate of contemporary receiver combined noise and resolution error is also included.

Satellite Geometry. As stated above, navigation solution accuracy is a function of UERE and satellite geometry. For example, assume a receiver's altitude and clock offset from GPS system time are known. The receiver's position can then be obtained by making ranging measurements to two satellites. **Figure 9** shows that the area of uncertainty in user position is minimized when the satellites form a right angle to one another. **Figure 9** indicates that the area of uncertainty increases as the angle formed by the satellites becomes greater than 90 degrees. The error contributed by satellite geometry to a user navigation solution is represented in terms of geometric dilution of precision (GDOP). GDOP is a function of the pointing vectors formed by the user-to-satellite geometric relationship. GDOP is a dimensionless factor that varies with satellite orbits, number of operational satellites, user position, and observation time. GDOP represents the square root of the trace (i.e., diagonal) of a covariance matrix containing estimates of pseudorange measurement errors in the receiver Kalman filter. The trace contains the variances of position errors in three dimensions and receiver clock offset error. In the representation below, it is assumed that all pseudorange errors are independent, zero mean, and have unity variance.

$$GDOP = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2} \qquad (16)$$

In this form, GDOP doesn't have simple physical meaning and can't be used to compute user position or time accuracy. Decomposing GDOP results in:

Position Dilution of Precision (PDOP)

$$PDOP = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \qquad (17)$$

Horizontal Dilution of Precision (HDOP)

$$HDOP = \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (18)$$

Vertical Dilution of Precision (VDOP)

$$VDOP = \sigma_y \qquad (19)$$

Time Dilution of Precision (TDOP)

$$TDOP = \sigma_t \qquad (20)$$

PDOP represents the contribution of satellite geometry to position error in three dimensions. The lateral and vertical contributions are represented by HDOP and VDOP, respectively. TDOP is a measure of error contributed by the satellite geometry incurred when determining the receiver clock bias error.

Computation of Measurement Accuracy. The product of the UERE and DOP factor yield measurement accuracy in a particular dimension. For example, the 2 drms value of horizontal measurement accuracy for a PPS receiver with an HDOP of 1.5 is computed as follows:

$$\begin{aligned} Horizontal\ error,\ 2drms(95\%) &= 2\ UERE\ HDOP \\ &= 2 * 6.6 * 1.5\ m \\ &= 19.8\ m \qquad (21) \end{aligned}$$

Therefore, approximately 95 percent of all measurements will fall within a 19.8 m radius circle centered about the user. VDOP is generally larger than HDOP, and on average may be approximated as:

$$VDOP \approx 1.73\ HDOP \qquad (22)$$

Therefore,

$$PDOP = \sqrt{VDOP^2 + HDOP^2} \qquad (23)$$

$$PDOP \approx \sqrt{HDOP^2 + HDOP^2} \qquad (24)$$

$$\approx 2HDOP$$

Using this approximation, PPS horizontal and vertical errors were calculated for PDOP values ranging from 2.5 to 6. These curves are illustrated in **Figure 10**. **Figure 11** denotes SPS accuracy for the same range of PDOP and the C/A-code UERE delineated in **Figure 7**.

Availability

"Availability is an indication of the ability of a system to provide usable service within the

specified coverage area."[2] For most GPS applications, availability is limited by accuracy (DOP criteria). The user needs to be aware of the maximum DOP anticipated for the duration of system usage (i.e., aircraft precision approach, vessel harbor navigation, land or marine surveying, etc.). For example, if the maximum horizontal error for a particular land survey using DGPS is 10 m, 95 percent, the maximum tolerable HDOP is:

Horizontal error, 2drms(95%) = 2 UERE HDOP

$$10 = 2 * 1.5 * HDOP_{max} \qquad (25)$$

Therefore,

$$HDOP_{max} = 3.33$$

DOP can be predicted by off-the-shelf computer programs. Using data obtained from any GPS bulletin board service, these programs compute all DOP types for a given location and time. The user must also take into account any failed satellites and/or scheduled maintenance actions. The U.S. Coast Guard provides GPS constellation data and status via a dial-up computer bulletin board at (703)313-5910.

Integrity

"Integrity is the ability of a system to provide timely warnings to users when the system should not be used for navigation."[2] A single GPS satellite illuminates approximately one third of the Earth; thus, any satellite failure will affect a large number of users. The OCS checks received monitoring station data for anomalies and issues warnings to users in the event a failure is detected. Integrity is enhanced by the use of differential GPS, DGPS, (described in the next section) and receiver autonomous integrity monitoring, RAIM. RAIM is a receiver-embedded algorithm that uses five or more satellites with good PDOP for fault detection. Other algorithms exist that incorporate measurements from at least six satellites to provide a fault detection and isolation capability.

Differential GPS (DGPS)

DGPS is a technique that offers improvements in GPS performance. DGPS enhances position accuracy by removing correlated errors between two or more receivers viewing the same satellites. Removal of correlated errors enables precise determination of the distance between receivers. Absolute or relative distance

measurements can be obtained. DGPS requires processing for correlated error extraction. In real-time applications, a data link capability is required. Error removal can use position, pseudorange, and/or carrier phase information.

Two popular techniques of implementing DGPS are local area DGPS (LADGPS) and wide area DGPS (WADGPS). LADGPS employs receivers located within line-of-sight (LOS) of each other. Correlated errors between two receivers can be removed by operating one GPS receiver (denoted as the reference receiver) at a surveyed location, comparing the electronically derived GPS position to the surveyed position, and broadcasting the difference in positions (i.e., error) to suitably equipped users within the area to correct their navigation solutions. This technique is generally not used because it restricts both receivers to view the same satellites or the reference receiver must compute corrections for all possible combinations of satellites in view. If eight satellites are in view and measurements to four or more satellites are required for a navigation solution, there are 163 possible combinations of four satellites or more to compute. This technique is generally prohibitive for real-time applications due to data latency, unless the user set is commanded to track the same set of satellites as the reference receiver.

Modern reference receivers compute the difference between the "true" reference receiver-to-satellite range (i.e., distance from the surveyed location to the ephemeris-reported position) and the pseudorange. Range differences (i.e., pseudorange measurement errors) for all satellites in view are transmitted via a communications link to users for incorporation in their pseudorange measurements, thereby improving user position accuracy. A ground-based integrity monitoring station assesses the quality of the correction broadcast before and during their transmissions and provides an alert if the correction broadcasts do not meet performance specifications. In general, existing PPS and SPS DGPS equipment performance is similar and can be represented by the estimated UERE budget delineated in **Figure 12**. LADGPS horizontal and vertical position errors are shown in **Figure 13**.

WADGPS is a means for improving user navigation sensor accuracy over a large area (e.g., CONUS and Canada, France, etc.). WADGPS is presently in development by the FAA and is comprised of local monitoring stations distributed across the large area, a central processing facility (CPF), a communications ground Earth station, and geostationary satellites. The system provides ephemeris and satellite clock bias error corrections for each GPS satellite, plus ionospheric time delay parame-

ters. Ephemeris and satellite clock errors, as well as ionospheric delay data, are generated by the ground monitoring stations and transmitted to the CPF via a terrestrial communications link. At the CPF, ephemeris and satellite clock corrections are computed and then forwarded to the ground Earth station for uplinking to geostationary satellites for rebroadcast to suitably equipped users within the satellite coverage area. The corrections are applied to the user receiver pseudoranges and collected ephemeris data in order to improve user navigation accuracy. It's envisioned that WADGPS accuracy performance will be an improvement to standalone GPS, but somewhat less than that provided by LADGPS.

The WADGPS network will also include an integrity failure and user warning capability. Integrity and WADGPS correction data will be generated by the ground monitoring network. One important current design goal is to provide an integrity failure warning to users via geostationary satellite within six seconds.

Applications

A multibillion dollar GPS market is projected. Applications range from leisure hiking to space shuttle landing guidance. Land uses include surveying, vehicle tracking, stolen vehicle recovery, public transportation, and fleet management. Combining GPS data with digitized maps permits a user to visualize her/his real-time position. There are also, several private entities broadcasting DGPS corrections over FM subcarriers. This service requires the purchase of an FM data receiver and is available for a monthly fee. The fee structure is based on the degree of accuracy the user requires.

GPS receivers are in wide use by the commercial/leisure marine community. Some manufacturers offer receivers that float. At present, the U.S. Coast Guard is establishing a coastal DGPS network that includes Alaska, most of CONUS, Hawaii, the Great Lakes, and the U.S. Caribbean Islands. Corrections are broadcast using the existing marine beacon system on approximately 300 kHz.

Current aviation usage encompasses enroute, terminal, and nonprecision approach applications. These phases of flight use standalone GPS. GPS applicability to precision approach is under investigation. The use of GPS with a communications link for automatic aircraft position reporting is expected to revolutionize air traffic control (ATC). This technique

is known as automatic dependent surveillance (ADS). At present, aircraft use INU and OMEGA for oceanic en route navigation. These positions are transmitted on an hourly basis via high frequency voice to ATC. ADS would automatically transmit an aircraft's GPS position to ATC approximately every five minutes over a digital data link, providing more accurate and timely position information.

GPS is also used in a multitude of military missions. Several military applications are listed below:

Rendezvous
All-weather air drop
Search and rescue
Close air support
Range Instrumentation
Weapons delivery
Target acquisition aid
Target reconnaissance
Missile guidance
Bomb scoring

Closing comments

GPS is here. The system is operational and has revolutionized the science of navigation by offering better positioning technology than ever. A person can now locate herself/himself nearly anywhere in the world with a relatively low-cost receiver. The system continues to prove its military utility as evidenced in Operation Desert Storm. New civilian and military applications are arising almost daily, and the number of civilian users is expected to outnumber those in the military. Within the next few years, GPS will influence some part of our daily lives. ■

REFERENCES
1. R. Langley. "The Mathematics of GPS," GPS World Magazine, July/August 1991.
2. DOD/DOT, 1992 Federal Radionavigation Plan, January 1993.
3. E. Maloney, Dutton's Navigation and Piloting, Naval Institute Press, 1978.
4. R. Keegan, The New Magnavox GPS Receiver Technology. Paper presented at Institute of Navigation Technical Meeting, September 1990.
5. NAVSTAR-GPS Joint Program Office, NAVSTAR GPS USER EQUIPMENT—INTRODUCTION. Department of Defense, February 1991.
6. K. Doucet and Y. Georgiadow, "The Issue of Selective Availability," GPS World Magazine. September/October 1990.
7. R. Milliken and C. Zoller, "Principle of Operation of NAVSTAR and System Characteristics," GPS Paper in Navigation, Volume 1. Institute of Navigation, 1980.
8. T. Logsdon, The NAVSTAR Global Positioning System, Van Nostrand Reinhold, 1992.
9. J. Spilker, Jr., "GPS Signal Structure and Performance Characteristics," GPS Papers in Navigation, Volume 1, Institute of Navigation, 1980.
10. J. Collins, et al., GPS Theory and Practice, Second Edition, Springer-Verlag, 1993.
11. Radio Technical Commission for Aeronautics (RTCA) Special Committee 159, Minimum Operational Standards for Airborne Supplemental Navigation Equipment Using GPS, RTCA, 1991.

David Barton, AF6S, sent us his "Yagi Gain Redux" in response to a letter from Roger Cox, WBØDGF.

Dear Editor:

Roger Cox, WB0DGF, noted an error in my article, "Yagi Gain Versus Boom Length," in the Winter '94 *Communications Quarterly*, in listing the HG215DX antenna's boom length as 8.4 wavelengths. He even supplied a copy of the page in the Hy-Gain manual that shows its correct value, 4.03 wavelengths. He also noted that the data in the article agrees well with the K1GQ formula:

$$G = 3Ln(B) + 9.85$$

where: G is free-space gain, dBd, and B is boom length in wavelengths

Figure 1 is a replot of the data, with the error corrected and a line for K1GQ's formula added. I deleted front-to-back ratio, since it isn't relevant to the subject, gain versus boom length. That allowed the vertical axis scale to be finer, for better readability.

The data seem to support the "quarter-wavelength step" idea—the observation that Yagi gain doesn't seem to improve much at first, one-quarter wavelength. The real-antenna data points fall below the formula line above about 0.25 wavelength by an average of roughly 1 dB

around 0.5 wavelengths, and don't get close to the formula values again until boom length is almost one wavelength.

Eyeballing the graph, it seems a somewhat steeper line would fit better above 0.5 wavelengths, but the formula would need to produce a "knee," and a lower slope below that boom length. The reprint Roger sent from the YO v.5 manual (page 21) also shows two other formulas, but all three have the same form; none make a "knee." An accurate gain-limiting formula may await someone's deeper understanding of the short-Yagi problem.

David Barton, AF6S
San Jose, California

Joe Orrico, WB6HRO, offers the following suggestion for those who lament the high cost of test equipment.

Dear Editor:

I have an idea for the "Technical Conversations" column that may be helpful to other amateurs.

The growing popularity of the 1.2-GHz band has made one thing very evident to repeater owners on this band: *the lack of test equipment.* There is a good selection of equipment covering this frequency range, but it will probably cost more than your house.

Since I already own a Cushman CE-6 service monitor, with an upper frequency limit of 1000
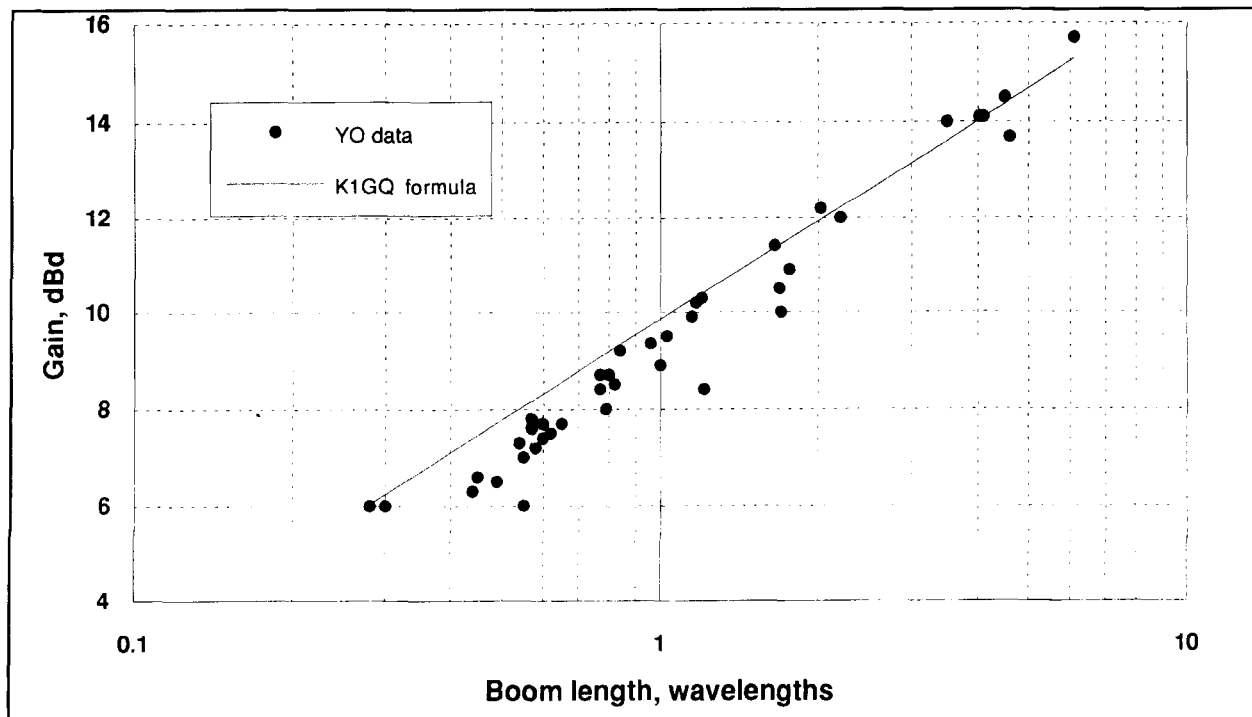
Figure 1. Revised Yagi gain versus boom length.

MHz, it became frustrating setting audio levels and deviation levels for the 1.2-GHz repeater in our linked system using a handheld radio.

Looking closer at the frequency generation scheme of the Cushman, it became apparent that all mixing was done between 1200 and 2200 MHz. Following that line, if frequencies under 1000 MHz are derived by taking the difference of the mixing frequencies, then frequencies of over 1000 MHz could be obtained by taking the sum of the frequencies.

As it turned out, if you want a frequency of 1280.00 MHz, you dial up 920.00 and a strong signal will be apparent on 1282.00 MHz. While the attenuator is no longer accurate at these frequencies, a stable phase locked loop signal and calibrated FM deviation is now available.

The transmitter output was then observed on the deviation scope with good sensitivity, using the internal Cushman Model 318A High Sensitivity Mixer, 20 to 1000 MHz. I believe that similar results would be evident using the other available preselector plug-in units.

It should be noted that signal generator output is higher on 1.2 GHz when the 900 MHz range is used, due to the filter and attenuator characteristics.

The formula to determine what service monitor dial frequency to use is:

1300 – f1 (1.2 GHz) + 900 = f2 (service monitor frequency)

For example:

1300 – 1282.050 + 900 = 917.950

1300 – 1294.500 + 900 = 905.500

I hope that others will find this information beneficial, as the older Cushman service monitors can be had at reasonable prices. Check with your local commercial shop as they might have a retired unit on a shelf.

Joe Orrico, WB6HRO
Whittier, California

A letter to Yardley Beers, W0JF, on his article "Double Resonant Antennas" (Winter 1994, page 57), led Yardley to write with information on the physical realization of networks.

Dear Editor:
With reference to my "Double Resonant Antenna" article,[1] Robert K. Zimmerman, NP4B, has sent me a letter which prompts me to write about the physical realization of networks—a very interesting but mathematically complex topic. My knowledge is superficial, hardly more than to let me add to the bibliography that he has supplied me.

The fundamental fact underlying this subject is the law that no signal can travel either through an electrical network or through bulk material with a speed faster than that of light through a vacuum. The speed in a network depends upon how its resistance and reactance vary with frequency. Nature is consistent by allowing the physical existence of only the networks whose reactance and resistance variations conform with this principle. These variations are described by some mathematical theorems. There are similar considerations concerning bulk matter, which are historically older than those for networks.

Specifically Zimmerman points out that my **Figure 7C**, a sketch of how the reactance of a center-fed wire antenna varies with frequency, has the form suggested by Foster's theorem.[2,3,4,5] This is one of the theorems to which I allude in the previous paragraph. This theorem pertains to arbitrary two-terminal networks containing only lossless lumped capacitors and inductors. However, with such a network, the number of resonances and antiresonances is finite and depends upon the complexity of the network. For the antenna, the number is infinite.

My justification for this sketch is an elementary one. In my earlier "ONETWORES" article,[6] I assume that the reactance looking into one half of a center-fed wire is that of a transmission line of characteristic impedance Z0, which, as is well known, varies as - cotangent of the angular length B. Therefore explicitly, the total reactance is:

$$X = -2\ Z0 \cot B = -2\ Z0/\tan B$$

My **Figure 7C** is just the sketch of -1/tan B versus B. In my calculations I have made the assumption that Z0 = 500 ohms, a constant. In practice Z0 is a slowly varying function of B, but the effect of including the variation does not change its general shape but only its numerical scale.

Foster's theorem, which concerns only reactance, is not the only network theorem which results from the requirement on the speed of signals. Bode[7] has shown that there is a partial dependence of the reactance variation upon the resistance variation of a network, whether two-terminal or otherwise. In general, whenever there is a rapid variation of resistance (or amplitude), there must be a rapid variation in reactance (or phase shift). The converse may or may not be true. It is possible to produce a phase shift of arbitrary size without much change in resistance, by introducing a transmission line. Thus Bode's theorem suggests a mathematical inequality rather than an equality, and it suggests the concepts of "minimum" and "non-minimum phase shift" networks.

Bode's theorem has at least one practical implication for amateur radio. It suggests a difficulty in using two unlike antennas as a phased array. It implies that whenever the SWR varies rapidly with frequency, there is a rapid change in the phase between the transmitter output and the radiated signal. Unlike antennas usually have unlike SWR curves. Therefore it is probable that optimum phasing can be established only over a very narrow band of frequencies.

The theoretical work on bulk matter centers upon what are called the "Kramers-Kronig Relations." For readers with an interest in this subject, I provide a partial bibliography in References 8, 9, and 10.

<div align="right">

Yardley Beers, W0JF
Boulder, Colorado

</div>

REFERENCES
1. Yardley Beers, W0JF, "Double Resonant Antennas with Loading Reactors." Communications Quarterly, Winter 1994, page 57.
2. R.M. Foster, "A Reactance Theorem," Bell System Technical Journal, 3, 259, April 1924.
3. F.E. Terman, Radio Engineers' Handbook, McGraw-Hill, New York, 1943, page 200.
4. R.H. Dicke in Principles of Microwave Circuits, Vol. 8 of the MIT Rad. Lab. Series, McGraw-Hill, New York, 1948, Sec. 5.24.
5. S. Ramo and J.R. Whinnery, Fields and Waves in Modern Radio, John Wiley and Sons, New York, 2nd edition, 1953, page 465.
6. Yardley Beers, W0JF, "ONETWORES: A C-64 Program for the Analysis of Single and Double Resonant Dipoles," Communications Quarterly, Spring 1993, page 91.
7. H.W. Bode, Network Analysis and Feedback Amplifier Design, Van Nostrand, 1945.
8. J.D. Jackson, Classical Electrodynamics, John Wiley and Sons, New York, 2nd edition. 1975, Secs. 7.10 and 7.11.
9. A. Yariv, Quantum Electrodynamics, John Wiley and Sons, New York, 1967, pages 131 and 455.
10. John S. Toll, Physics Review 104, 1760, 1956.

After reading the letter from Mason Logan, K4MT's letter regarding the transformerless noise bridge, Alfred Popodi, AA3K/OE2APM had some comments.

Dear Editor:

I refer to the contribution by Mason A. Logan, K4MT ("Technical Conversations," Fall 1993, page 6), describing a transformerless noise bridge.

I have built this circuit and have found a prohibitively large error when measuring large impedances at high frequencies. For example, a 180-ohm resistor measures about 180 ohms at 1.8 MHz, but only 166 ohms at 28 MHz. If I connect an additional 27-pF capacitor between the two shielded boxes, this reading drops from 166 to 127 ohms.

The reason is that the entire interbox capacitance, C_{ib}, does not lie in parallel to bridge point C, because of the unavoidable parasitic impedance, Z_2, that exists between point C and the connecting point C' at the inner box. The enclosed schematic (**Figure 1**) shows the actual circuit configuration. This circuit is very sensitive to any impedance between the points C and C'. If you add, for example, a 10-ohm resistor in this line, at point C, the reading drops from 166 to 137 ohms. Because it is practically impossible to avoid this parasitic impedance, because of mechanical constraints, it also impossible to avoid this error.

However, if you add a variable capacitor, C_{bal}, between the noise generator output and ground, this error can be completely eliminated. This capacitor restores the bridge balance, as seen from the figure. The interbox capacitance, C_{ib}, is separated from bridge point C by the small parasitic impedance Z_2. Also, on the A side, there is a parasitic impedance, but it is less important.

Despite the mechanical difficulties in building this bridge, the effort is worthwhile, because its accuracy is better than that of a transformer bridge, as my tests have shown. However, I have only measured purely resistive loads. In addition, the balancing control is an excellent means for optimizing bridge performance, which other bridges do not have.

<div align="right">

Alfred E. Popodi, AA3K/OE2APM
Salzburg, Austria

</div>
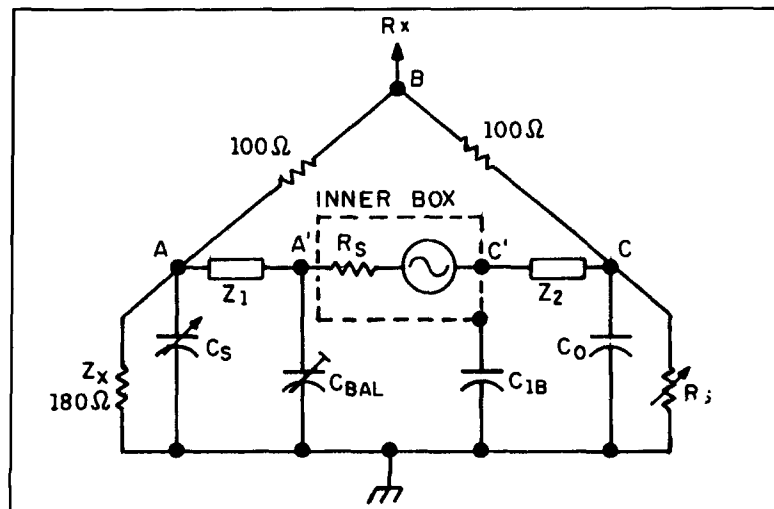
Figure 1. Schematic of the actual circuit configuration.

Glenn E. Prescott, *WBØSKX*
University of Kansas
Lawrence, Kansas,
and
Phil Anderson, *WØXI*
Kantronics, Inc.
Lawrence, Kansas

HYBRID ARQ FOR HF DATA TRANSMISSION

Forward error correction on demand

The G-TOR data communications protocol was recently introduced by Kantronics Inc. as an inexpensive means of improved digital communications in the HF amateur radio frequency bands. Although many modern digital communications innovations have been incorporated into G-TOR, the single most essential element in this protocol is the use of a hybrid ARQ scheme in combination with an invertible, half-rate forward error correcting code. In this article, we will describe several types of automatic repeat query (ARQ) systems in use today, and contrast these with protocols based on the use of forward error correction (FEC). In the process, we will see why hybrid ARQ—a combination of FEC and ARQ, such as that used in the new G-TOR system—is ideal for communication on the HF bands. We will also provide a detailed description of the encoding and decoding procedure for the Golay forward error correcting code in order to give G-TOR users a better understanding of the inner workings of the protocol.

ARQ fundamentals

There are essentially two techniques for controlling transmission errors in data communication systems: ARQ and FEC. The ARQ system uses an error detection code to detect the presence of errors in a received block of data, and a return link to request retransmission of any blocks found to be in error. A message is encoded using a relatively simple, high-rate code. The *rate* of a code, R_c, is simply the ratio of the number of bits entering the encoder to the number of bits leaving the encoder during any time interval. A high-rate code is one that appends relatively few parity bits to the information sequence in order to detect errors at the receiver. **Figure 1** illustrates the encoder for the Golay code in which the input bit stream is divided into k-bit *information words* (k = 12 in this case) and fed into the encoder. In the encoding process, 12 parity bits are added to the input information word to produce an n-bit *codeword* (in this case, n = 24). The code rate is defined as $R_c = k/n$ or, in this case, 12/24 or 1/2. The code rate is always a number less than 1.

At the information receiving station (IRS), parity checking is performed on the received codeword. If the parity check reveals that the received codeword contains no errors (i.e., it is a valid codeword), the information is recovered and delivered to the user or to a buffer to await further processing. At the same time, a positive acknowledgment (ACK) is transmitted across the return channel to the information sending station (ISS). On the other hand, when there is a parity failure on any codeword (i.e., the received codeword is not a valid codeword), a negative acknowledgment (NACK) is transmitted to the information sending station. Upon
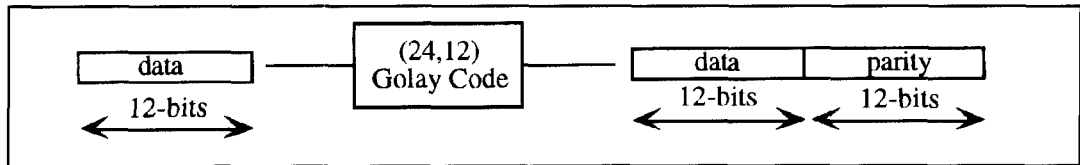
Figure 1. The (24,12) rate 1/2 Golay encoder.
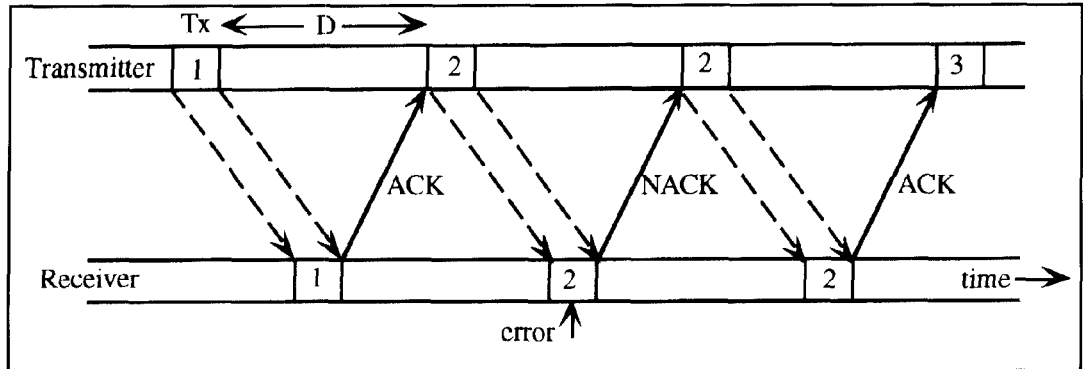
Figure 2. Stop-and-wait ARQ.

receipt of the NACK, the information sending station retransmits the codeword. Retransmission continues until the codeword is successfully received. The error detection can be accomplished on a word-by-word or on a frame-by-frame basis. In amateur communications protocols like AX.25, groups of codewords (i.e., frames) are normally protected by a single overall error detection code—the CRC. Using one of these error detection techniques, erroneous data is delivered to the user only if the receiver fails to detect the presence of errors. The ARQ scheme is used in data communication systems because it is simple and provides high system reliability. However, ARQ systems have one severe drawback: their throughput efficiency falls rapidly with increasing channel error rate. Throughput efficiency is defined as the ratio of the rate of the data bits successfully received to the data rate possible when data is transmitted continuously (and successfully) without coding.

Stop-and-wait ARQ

There are three variations of ARQ in popular use. The simplest and most popular system is *stop-and-wait* ARQ, in which the ISS transmits a block of data (one or more codewords) and waits for an ACK from the IRS. Further data transfer across the channel does not proceed until the present block of data has been received correctly. The AX.25 protocol used in HF and VHF packet radio employs a stop-and-wait ARQ. G-TOR employs a variation of this system, which will be described later. The stop-and-wait ARQ system is especially effective

for use with half duplex channels (used extensively in amateur radio) in which the propagation delay is relatively short. A typical stop-and-wait ARQ transmission sequence is shown in **Figure 2**. Obviously, the ACK/NACK transmissions are also of concern because incorrect receipt of either of these will cause the data to be retransmitted, reducing throughput efficiency. Because they contribute only to overhead, the ACK or NACK transmissions are usually very short. In G-TOR, the ACK/NACK transmissions are 16 bits long and can tolerate the presence of several errors. They are always transmitted at the slowest (100 baud) data rate.

Go-back-N ARQ

When the round trip delay (D) from the transmitter to the receiver and back is very long, the transmitter spends an excessive amount of time waiting for the ACK or NACK, which reduces the throughput efficiency. In this situation, the *go-back-N* ARQ approach may be preferred because it increases the throughput efficiency by continuing the transmission of code blocks without waiting for the return acknowledgment for each one. When a block is received error-free, an ACK is returned to the ISS. As long as the ISS continues to receive ACKs, it will continue to transmit new blocks. When the receiver detects an error, it returns a NACK and ignores the next N received blocks. Upon receiving a NACK, the ISS backs up N blocks in the sequence and transmits the last N blocks. The go-back-N ARQ sequence is illustrated in **Figure 3**. When the round trip delay is large and the error rate is small, this technique
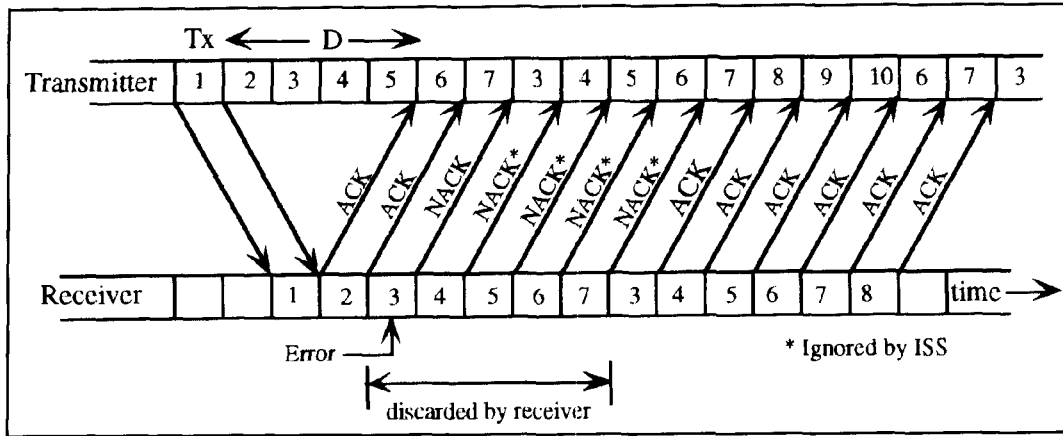
Figure 3. Go-back-N ARQ.

has considerably higher throughput efficiency than the stop-and-wait approach. However, as the error rate becomes higher, the throughput efficiency decreases significantly.

A modified version of the go-back-N ARQ system is employed in AX.25. It is included in most TNCs as a feature that permits up to 7 frames to be concatenated and transmitted in a single burst. In this case, if one of the frames is detected in error, the IRS will return a NACK to the ISS indicating the number of the failed frame, and the ISS retransmits the failed frame and all succeeding frames in the extended packet. This concept is illustrated in **Figure 4**.

Selective-repeat ARQ

The *selective repeat* ARQ is the third important technique for ARQ error control. This procedure, illustrated in **Figure 5**, is not currently used in amateur HF digital communications. When the IRS detects a block with errors, it returns a NACK including an identification of the block that is in error. The transmitter continues to send new blocks and repeats only those blocks for which a NACK has been received. This technique achieves higher throughput efficiency than the other two by

repeating only those blocks that are received in error. This approach also makes the throughput efficiency of the selective repeat ARQ independent of round trip delay. However, the additional processing required to identify the block in error and allocate buffer space for the received data until the bad block is replaced, makes the selective repeat ARQ procedure more complex to implement.

Of the three techniques, the stop and wait is the least complex and the selective repeat is the most complex. Selective repeat attains the highest throughput efficiency, stop-and-wait the lowest. The throughput efficiencies of all three systems are reduced when the channel error rate is high. It is not unusual for the throughput to drop to unusable levels at high error rates because the transmitter spends most of its time repeating code groups. As the round trip delay increases, the throughput efficiencies of both the stop-and-wait and go-back-N systems decrease, while the efficiency of the selective-repeat unaffected by the length of the round trip delay.

ARQ versus FEC

Although the conventional ARQ schemes are widely used in data communications, they have

Figure 4. Go-back-N ARQ (as used in AX.25).

Figure 5. Selective repeat ARQ.

$$G = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
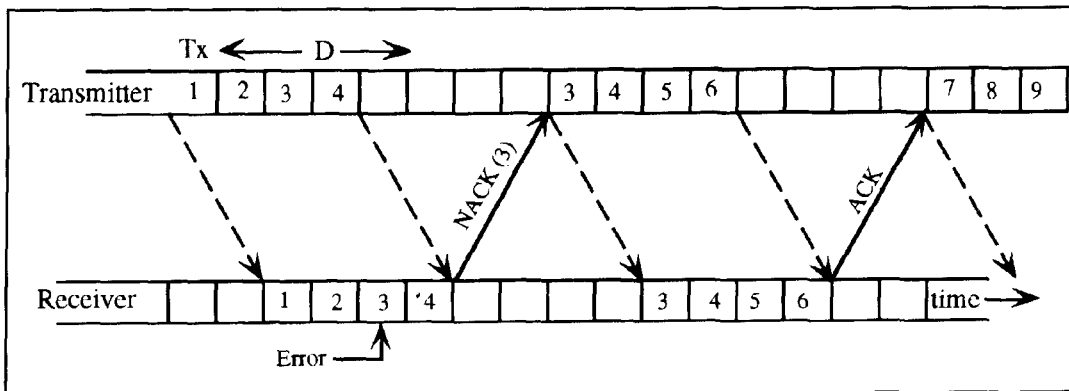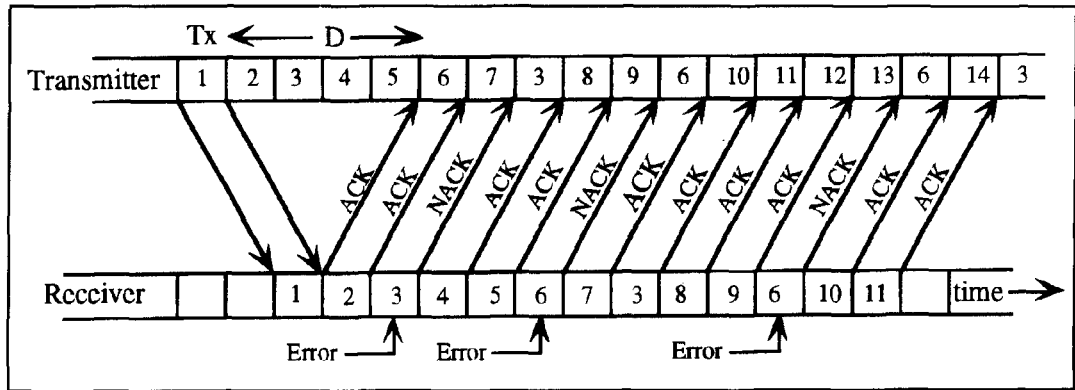0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}$$

Figure 6. Generator matrix for the (24,12) Golay code.

one major drawback: their throughput efficiency falls rapidly with increasing channel error rate. That brings us to consider FEC-only techniques, in which an error correcting code is used to correct errors in the transmitted data. In the FEC system, when the receiver detects errors in the received transmission, it attempts to determine the error locations and then corrects the errors. If the IRS fails to determine the exact error locations, the received data block will be incorrectly decoded and erroneous data will be delivered to the user.

No retransmission is required in an FEC system. Therefore, the throughput efficiency of the system remains constant regardless of the channel error rate. However, FEC systems have some significant drawbacks. First, when a received codeword is detected in error, it must be decoded, and the decoded message must be delivered to the user regardless of whether it is correct or incorrect. Because the probability of a decoding error is usually much greater than the probability of an undetected error, it is hard

to achieve high system reliability using only FEC. Second, in order to obtain high system reliability, a long and powerful code must be used. The longer the code (large *n*), the lower the throughput efficiency, and the more complex the decoding procedure due to a larger collection of error patterns that must be recognized and corrected. This makes decoding expensive and hard to implement. For this reason, ARQ is usually preferred over FEC in data communication applications.

Hybrid ARQ

The limitations of both ARQ and FEC systems can be overcome somewhat if the two error control schemes are properly combined. Such a combination of the two basic error control schemes is referred to as *hybrid* ARQ. A hybrid ARQ system consists of an FEC subsystem contained within an ARQ system. The FEC portion keeps retransmissions at a minimum, so

as the channel begins to degrade, the through-put efficiency is kept constant. As the channel error rate degrades further, and the error correction capability of the FEC is exceeded, the ARQ system is used to request retransmission of the blocks that cannot be corrected. Therefore, a proper combination of FEC and ARQ can provide higher reliability than an FEC system alone and higher throughput than the system with ARQ only.

There are two popular error detection techniques employed in hybrid ARQ systems. The first approach is to use an error correcting code designed for simultaneous error correction and detection. Error correcting codes have the ability to detect more errors than they can correct—a result of the mathematical characteristics of all such codes. For example, the Golay code employed in G-TOR can correct up to 3 errors in a 24-bit sequence (as we will describe later), while the same code can detect as many as 7 errors in the same sequence. The second approach is to use a cyclic redundancy code (CRC) to detect errors in each block of data. When this approach is used in a Hybrid ARQ scheme, the CRC code allows the IRS to detect errors in the received block of data, while the FEC code is used exclusively for error correction. There are several CRC codes which are suitable for use; for example (see **Reference 1**). Both AX.25 Packet and G-TOR use the same CRC code, which produces a 16-bit check sum. This error detection code can detect all single and double errors, all errors with an odd number of bits, all burst errors of length 16 or less, and 99.999 percent of 17-bit and longer bursts.

Type-I hybrid ARQ

There are type-I and type-II hybrid ARQ systems. Type-I hybrid ARQ systems are characterized by the simultaneous use of error correction and error detection in each block of data. When the received block is determined to be in error, the IRS first attempts to correct the errors. If the number of errors present in the block of data is within the error correcting capability of the code, the errors will be corrected and the data will be passed to the user. If the received data contains a large number of errors that exceeds the error correcting capability of the code, an uncorrectable error pattern will be detected. In this case, the IRS detects the presence of errors even after the error correction procedure has been attempted, and requests a retransmission from the ISS.

Obviously, this type of system requires more parity check bits than a system that uses only error detection. Therefore, the throughput efficiency is smaller than conventional ARQ schemes. However, when the channel is poor

$$p_1 = c_1 \oplus c_2 \oplus c_3 \oplus c_4 \oplus c_5 \oplus c_8 \oplus c_{11}$$

$$p_2 = c_2 \oplus c_3 \oplus c_4 \oplus c_5 \oplus c_6 \oplus c_9 \oplus c_{12}$$

$$p_3 = c_1 \oplus c_2 \oplus c_6 \oplus c_7 \oplus c_8 \oplus c_{10} \oplus c_{11}$$

$$p_4 = c_2 \oplus c_3 \oplus c_7 \oplus c_8 \oplus c_9 \oplus c_{11} \oplus c_{12}$$

$$p_5 = c_1 \oplus c_2 \oplus c_5 \oplus c_9 \oplus c_{10} \oplus c_{11} \oplus c_{12}$$

$$p_6 = c_1 \oplus c_4 \oplus c_5 \oplus c_6 \oplus c_8 \oplus c_{10} \oplus c_{12}$$

$$p_7 = c_1 \oplus c_3 \oplus c_4 \oplus c_6 \oplus c_7 \oplus c_8 \oplus c_9$$

$$p_8 = c_2 \oplus c_4 \oplus c_5 \oplus c_7 \oplus c_8 \oplus c_9 \oplus c_{10}$$

$$p_9 = c_3 \oplus c_5 \oplus c_6 \oplus c_8 \oplus c_9 \oplus c_{10} \oplus c_{11}$$

$$p_{10} = c_4 \oplus c_6 \oplus c_7 \oplus c_9 \oplus c_{10} \oplus c_{11} \oplus c_{12}$$

$$p_{11} = c_1 \oplus c_2 \oplus c_3 \oplus c_4 \oplus c_7 \oplus c_{10} \oplus c_{12}$$

$$p_{12} = c_1 \oplus c_3 \oplus c_5 \oplus c_6 \oplus c_7 \oplus c_{11} \oplus c_{12}$$

Figure 7. Parity generation equations for the 24,12 Golay code.

and the error rate is high, the type-I hybrid ARQ system provides higher throughput than a conventional ARQ scheme because its error correction capability will reduce the retransmission frequency. CLOVER is a popular system that apparently uses type-I hybrid ARQ.[2] More correctly, we can say that CLOVER uses a type-I hybrid stop-and-wait ARQ system, because stop-and-wait ARQ is the underlying procedure. CLOVER employs one of several possible Reed-Solomon error correction codes, depending upon the block length selected. However, when the error correcting ability of the Reed-Solomon code is exceeded, a retransmission request is sent to the ISS.

Type-II hybrid ARQ

In the type-II hybrid ARQ system, forward error correction is employed only when it is needed. This is accomplished by storing the parity bits at the ISS and transmitting the data along with the CRC. When the data is received at the IRS, the CRC check sum is computed to determine if the block was received correctly. If the check sum indicates that errors are present in the received block, a NACK is returned to the ISS. Receipt of the NACK prompts the ISS to transmit the parity bits, which are then used by the IRS to correct errors in the original block of data. There are practical reasons for requiring that the error correction code used in this scheme be a 1/2-rate code. For example, with a 1/2-rate code, there is one parity bit for each information bit, which permits the forma-

---

Decoding Algorithm for the (24,12) Binary Extended Golay Code [6]

let $G = [I_{12}B]$ be the generator matrix.
let $I_1, I_2, ..., I_{12}$ be the columns of B and $r_1, r_2, ..., r_{12}$ the rows.
Let 0 denote the 12-tuple of all zeros.
Let $x^{(i)}$, $y^{(i)}$ denote binary 12-tuples in which only component i is non-zero.
Let r be a vector received by the decoder.
Let $w(s)$ denote the hamming weight of some vector s.

❶ Compute $s = Gr^T$

❷ if the $w(s) \leq 3$,
  then set $e = (s^T, 0)$ and go to ❽

❸ if $w(s + I_i) \leq 2$ for some I_i, $1 \leq i \leq 12$,
  then set $e = [(s + I_i)^T, y^{(i)}$ and go to ❽

❹ Compute B^Ts.

❺ if $w(B^Ts) \leq 3$,
  then set $e = [0, (B^Ts)^T)$ and go to ❽

❻ If $w(B^Ts + r^T_i) \leq 2$ for some r_i, $1 \leq i \leq 12$,
  then set $e = [x^{(i)}, (B^Ts)^T + r_i]$ and go to ❽

❼ At least 4 errors have occurred. Retransmit, and restart at ❶

❽ Decode r to $r - e = c$.

Figure 8. Decoding algorithm for Kasami decoder.

tion of a parity check frame of the same length as the original data frame. If a 1/2-rate code were used in the type-I hybrid ARQ, it would cause a 50 percent decrease in throughput efficiency at all times. However, by using forward error correction only when needed, the type-II hybrid ARQ scheme allows the effective use of powerful error correcting codes in the most efficient means possible.

A further refinement to the type-II hybrid ARQ system can be realized by using an *invertible* 1/2-rate forward error correcting code. The invertible property means that the original information bits can be uniquely recovered from the parity bits. The advantages of this approach can be illustrated by an example. Before transmitting a block of data to the receiver, the ISS first applies the CRC to the data for error detection, and then encodes the data using the 1/2-rate FEC. The parity bits generated by this process are saved at the transmitter and the information bits (including the CRC) are sent. When the IRS receives the data, the check sum is computed. If the check sum passes, the data is accepted and sent to the user, and an ACK is returned to the transmitter. If the check sum fails, a NACK is returned to the transmitter. When the transmitter (ISS) receives

an ACK, the parity bits from the first data block are discarded and the next block of data is processed. If a NACK is received, the transmitter sends the block of parity bits which had been held back. When the block of parity bits arrives at the receiver (IRS), the first action taken is to invert the parity bits to obtain data bits. Once data bits are obtained the check sum is computed. If the data (which was obtained by inverting the parity bits) passes the check, it is accepted and sent to the user. If it fails the check, then the parity bits are combined with the data bits from the first transmission and processed by the FEC decoder. In this way, *two* transmissions provide the receiver with *three* opportunities to obtain an error-free block of data.

G-TOR uses the type-II hybrid stop-and-wait ARQ system in combination with a 1/2-rate invertible Golay code for error correction and a 16-bit CRC code for error detection. The CRC code is the same one used by the AX.25 packet radio protocol, and recommended by the CCITT.

Memory ARQ

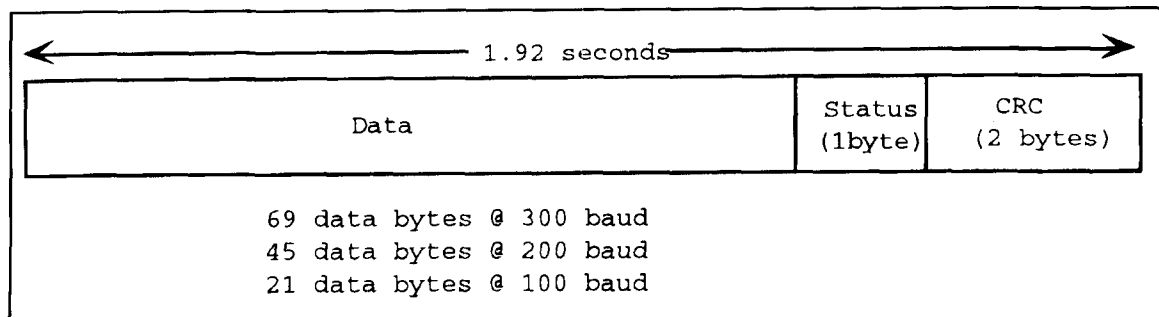Memory ARQ (MARQ) can also be considered a hybrid form of one of the conventional

```
←──────────────── 1.92 seconds ────────────────→

┌──────────────────────────────────────┬─────────┬───────────┐
│                                       │ Status  │    CRC    │
│               Data                    │ (1byte) │ (2 bytes) │
│                                       │         │           │
└──────────────────────────────────────┴─────────┴───────────┘

            69 data bytes @ 300 baud
            45 data bytes @ 200 baud
            21 data bytes @ 100 baud
```

Figure 9. G-Tor frame structure before interleaving.

ARQ techniques, since it can be used with any one of them. In fact, MARQ can be used easily with both of the hybrid forms described previously. This procedure is sometimes referred to as *ARQ-with-memory*.[3] The basic idea behind MARQ is that a received block of data which is found to contain errors should not be discarded because it contains *some* amount of useful information about the transmitted block. If the retransmission of this block is also in error, the collective information present in the first block and the retransmission can be used to correct certain errors in the two blocks to obtain a single error-free block of data. If correction fails using two blocks, additional retransmissions can be used until enough redundancy is present to allow correction. Increasing the redundancy through retransmissions is the key to this approach.

There are two ways to increase the redundancy, using either digital or analog techniques. The simplest approach is to use majority decoding of the received data block. As long as an odd number of retransmitted blocks have been received, select every bit according to the majority vote. For the analog approach, a copy of the received data signal is saved at the receiver before it is converted to a bit stream. If the bit stream fails the check sum, the stored signal is averaged with the retransmitted signal and this signal is converted into a bit stream and checked for errors. Many retransmissions can be averaged to improve the signal-to-noise ratio of the signal until an error-free block can be recovered. The drawback is that the complexity and cost of the analog system can be high because analog signals must be stored at the receiver. However, it is possible to store the signals in limited precision form using digital signal processing techniques. MARQ works especially well in a benign white noise environment, although a large number of retransmissions may be required to obtain an error-free data block. MARQ fares less well in the presence of interference and burst-noise. It is most effective when used in combination with the hybrid ARQ systems previously described.

MARQ is currently used in the PacTOR protocol for HF data transmission. An 8-bit analog-

to-digital converter is used to convert the received signal to a sequence of numbers representing the signal strength of each bit in the data block. If a block is found to have errors, retransmissions are similarly A/D converted and averaged with the previously stored signals.

Forward error correction and the Golay code

The Golay code belongs to a class of error correcting codes known as linear block codes. Block codes are simply fixed length codes—*k* information bits are fed into the encoder and *n* resulting code bits are fed out the other end. Of these *n* code bits, *n–k* are parity bits whose sole purpose is to correct any errors that may occur in the *n* bit codeword after traversing the channel. When selecting a code, it is important to remember that the code's error correcting capability lies in its distance properties. We say that the distance (more properly called the Hamming distance) between two binary code words is the number of places in which the code words differ. For example, consider two binary words:

$a = (0\ 1\ 1\ 0\ 1\ 0\ 1\ 0)$
$b = (0\ 1\ 1\ 0\ 0\ 0\ 0\ 1)$

These two words are separated by a distance of 3 because they differ in three bit positions. The error correcting capability of a code is dependent upon a single parameter—the minimum distance, d_{min}, of the code. The minimum distance of a code is the smallest Hamming distance between any two codewords in the entire code set. The error correcting capability of a linear block code, t, is related to the minimum distance of the code, d_{min}, by the following:

$$d_{min} = 2t + 1$$

We will forgo a detailed explanation of forward error correcting codes, and refer the interested reader to **Reference 4** for additional technical information on the design and analysis of

these codes. A recent article in *Communications Quarterly* provides a fundamental and intuitive understanding of linear block codes, especially Hamming codes.[5]

The Golay code

The original Golay code appends 11 parity bits onto 12 information bits; therefore, it is designated as a (23,12) code. Since its discovery by Golay in 1949, it has become the subject of study by many coding theorists and mathematicians. It has a minimum distance of 7 and is therefore capable of correcting any combination of three or fewer random errors in a block of 23 bits. This code is closely related to the popular Hamming codes in its mathematical characteristics. However, we are more interested in a modified version of the original Golay code that is far more powerful for our application than the original code.

The forward error correcting code used in G-TOR is characterized as an *extended binary Golay code*. It has numerous useful properties that have led to its use in communications and telemetry over the past 20 years. Its most notable application was in NASA's Voyager space mission, where it was used to transmit color images of Jupiter and Saturn in the early 80s. We decided to use the Golay code for the G-TOR protocol because of its simplicity and its powerful mathematical properties—most notably, the fact that it is *invertible*.

The (24,12) extended binary Golay code is formed by adding one parity bit to the original (23,12) Golay code. The added parity bit increases the minimum distance from 7 to 8 and produces a rate 1/2 code, which is far more useful than the rate 12/23 original Golay code. Since $d_{min} = 8$ for the extended Golay code, it can correct all triple errors, and some—but not all—four-error patterns. Because only about 19 percent of the four-error patterns can be corrected, the decoder is usually designed to correct only three-error patterns.

Encoding and decoding with the Golay Code

The Golay code can be implemented as either a block or a cyclic code. To implement the block code, we use the matrix multiplication formula $X = U \, G$, where U is the 12-bit information word, G is a 12 x 24 generator matrix, and X is the resultant 24-bit message (information plus parity). Encoding is performed using the G matrix by summing (modulo-2) the rows of G for which the corresponding information bit is a "1". The generator matrix used in G-TOR is shown in **Figure 6**. Note that this matrix is in

systematic form; that is, use of this generator matrix will result in code words that contain the information bits in the first 12-bit positions. The approach used in G-TOR is to generate the parity bits with a set of parity equations obtained directly from the generator matrix. The parity equations for the Golay code are shown in **Figure 7**. This set of equations is used to encode the information at the transmitter and is also used to invert the received parity block at the receiver when a retransmission is required.

Decoding the Golay code is not quite as simple as the decoding process for Hamming codes. Through years of research on decoding techniques, mathematicians have developed numerous techniques for decoding Golay codes. In order to remove errors from any 24-bit received Golay code word, each error must be precisely located. Once it is located, the state of the bit in error is simply toggled—a 0 is toggled to a 1, and vice versa. One of the simplest decoding algorithms for the Golay code is the Kasami decoder. We will not describe the operation of the Kasami decoder here, as **Reference 4** discusses it in detail. However, for the reader interested in implementing the decoder, we have provided the algorithm in **Figure 8**. From this procedure, a computer program can be written to correct errors in any received 24-bit Golay code words. Taken together, **Figure 7** and **Figure 8** form the heart of the G-TOR protocol.

A brief description of G-TOR

Now that we have established several fundamental principles concerning ARQ, FEC and Golay codes, we will show how these are integrated into the G-TOR protocol. The primary objective of the G-TOR protocol is to provide an improved framework for a structured interchange of digital data between stations operating in the HF amateur bands. The notable features of this protocol are summarized below:

• Golay forward error correction coding and full-frame data interleaving
• Huffman data compression and run-length encoding
• CRC error detection with hybrid ARQ
• Link-quality based baud rate: Binary FSK at 300, 200, or 100 baud
• Fault tolerant ACK/NACK

The primary benefit of these innovations is increased throughput; i.e., more bits communicated in a shorter amount of time. The real power of G-TOR resides in the properties of the (24,12) extended Golay error correcting code which provides increased resistance to interference and noise, and greatly reduces multipath induced data errors. To help the error

correcting code alleviate the effects of burst noise and interfering transmissions, each transmitted frame is interleaved to the maximum depth allowed by the frame size.

An important feature of the G-TOR protocol is that it uses a type-II hybrid stop-and-wait ARQ approach for error control. The error detection code transmitted with each frame is a two-byte cyclic redundancy check (CRC) code—the same one used in the AX.25 protocol. The CRC code is used to determine if the packet was received correctly before error correction is initiated. It is also used after error correction has been completed to ensure that the error correction process has successfully removed all errors in the packet.

One of the primary causes of reduced throughput on synchronous ARQ links is errors in the acknowledgment (ACK) signal. To help reduce needless retransmissions due to faulty ACKs, G-TOR uses fault tolerant ACKs and NACKs which allow the ISS station to tolerate a small number of errors in the feedback signal. This fault tolerant ACK concept allows the protocol to tolerate up to 3 bit errors within a received ACK frame. All of these features work together in an efficient manner to increase the throughput of G-TOR.

G-TOR employs a binary FSK transmission mode, allowing data to be transmitted at three different rates (300, 200, or 100 bits/sec) depending upon channel conditions. The lowest speed, 100 baud data transmission, is used for initiating contacts and for ACK/NACK transmissions. The G-TOR waveform consists of two phase-continuous tones (BFSK) whose spacing is specified at 200 Hz (mark = 1600 Hz, space = 1800 Hz); however, the system can still operate at the familiar 170 Hz shift (mark = 2125 Hz, space = 2295 Hz), or with any other convenient tone pairs.

The Huffman data compression code is used to remove redundancy from source data in such a way that the original data can be completely recovered or restored. Therefore, fewer bits will be traversing the channel to convey any given message. Use of data compression codes will increase data throughput and decrease the required transmission time. Both of these are valuable features to use on the HF channel. The G-TOR protocol uses Huffman coding during normal text transmission. However, when the data transmitted will not benefit from Huffman coding, the encoding process will be disabled by the transmitting station. The decision to use (or not use) Huffman coding is made on a frame by frame basis by the information sending station.

The G-TOR frame structure

The frame structure for a typical G-TOR data

frame (before interleaving) is shown in **Figure 9**. The data frame is 1.92 seconds in duration. Each data frame is composed of 72 bytes (at 300 baud), 48 bytes (at 200 baud), or 24 bytes (at 100 baud), depending upon channel conditions. A single byte near the end of the frame is devoted to command and status functions. The basic frame structure for the G-TOR protocol is based on the use of multiple 24-bit (triple-byte) words in order to be compatible with the Golay encoder. It is obvious from this diagram that G-TOR is similar to PacTOR and AMTOR in that it permits short duration synchronous transmissions. Synchronous operation is used to increase the throughput of the system in the presence of multipath fading, and to keep the number of overhead bits required to a minimum. No synchronization headers are used in this protocol. Synchronization is performed using the received data and precise timing.

The ACK codes are composed of multiple cyclic shifts of a single 15-bit pseudo-random noise (PN) sequence (an extra '0' bit is appended to the sequence for balance, so the total ACK word length is 16). A pseudo-random noise sequence is used for the ACKs because PN sequences have powerful mathematical correlation and distance properties that facilitate the identification of the appropriate ACK code, even in the presence of noise and interference.

G-TOR hybrid ARQ

The half-rate invertible Golay code provides an interesting dimension to the hybrid-ARQ procedure. After the message to be sent is processed through the Golay encoder at the transmitter, the message (data) bits are separated from the parity bits and two complete frames are formed. One frame consists of data bits and the other consists of parity bits. Keep in mind that the receiver can recover the message bits from the parity bits alone, so either frame can be used to transmit the data, as long as the receiver knows which to expect. *The procedure used in G-TOR is to alternate data frames and parity frames.* For example, we might expect the following sequence of frames to be transmitted when no errors are detected at the receiver:

D1 P2 D3 P4 D5 P6

where D1 is the first frame transmitted (a data frame), and P2 is the second frame transmitted (a parity frame), etc.

When an error is introduced into a frame because of noise or interference, the receiver will detect the error and request a retransmission. When the sending station receives the request, the complementary portion of the

frame (either data or parity) is sent. For example, the sequence illustrated below demonstrates a situation where two retransmissions are sent:

D1 P2 D3 P3 D4 P5 D6 P7 D7 P8

Retransmissions occurred for Frame #3 and Frame #7. In the case of Frame #3, one or more data bits were received in error and a retransmission was requested. Instead of retransmitting D3, the corresponding parity frame P3 is sent. When P3 arrives at the receiver, the first action that occurs after deinterleaving is to convert the parity bits to data bits and check for errors. If no errors were detected then the data is accepted and a new frame (D4) is requested. If P3 fails the CRC check, then this frame in its original form (i.e., parity bits) is combined with D3 and the combination (D3P3) is decoded by the Golay decoder to correct the errors that occurred in both D3 and P3.

In the case of Frame #7, the parity bits were sent first in P7. After deinterleaving the frame, converting the parity bits to data bits and performing the CRC check, the presence of errors prompts the receiving station to request a retransmission. Because the parity frame has already been sent, the sending station transmits the data frame, D7. Upon receipt, the data frame is deinterleaved and checked for errors. If no errors are detected, the data bits are accepted and a new frame (P8) is requested. If errors are detected in the data, the parity bits are combined (D7P7) and decoded by the

Golay decoder. Using this scheme, with two transmissions, the receiver can make three independent attempts to correct any errors that may occur in any frame. If this process still fails to produce an error-free frame, a retransmission is requested.

Conclusions

We have presented a review of six popular ARQ techniques and shown how these techniques are currently being used for data communications on the amateur HF bands. In addition, we provided a description of the new G-TOR system and demonstrated how it uses hybrid ARQ techniques to control errors induced by noise, interference and fading. We also discussed the Golay forward error correcting code and showed how to implement both the encoding and decoding algorithms. Finally, we concluded with an example to illustrate how G-TOR uses the Golay code within the hybrid ARQ framework to exchange data over HF radio. ■

REFERENCES
1. Andrew Tannenbaum, *Computer Networks*, 2nd Ed., Prentice Hall, Englewood Cliffs, NJ, 1988, pages 210–212.
2. Bill Henry and Ray Petit, "HF Radio Data Communication: CW to Clover," *Communications Quarterly*, Spring 1992, pages 11–24.
3. Pradeep Sindhu, "Re-transmission Error Control with Memory," *IEEE Transactions on Communications*, Com-25, No. 5, pages 473–479, May 1977.
4. Shu Lin and Daniel Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1983.
5. Lewis McIntyre, "Error Correction in Data Transmission," *Communications Quarterly*, November 1990, pages 88–94.
6. Scott Vanstone and Paul Oorschot, *An Introduction to Error Correcting Codes with Applications*, Kluwer Academic Publishers, Boston, 1989.

David Bowman, *GØMRF*
Reprinted with
permission from
Radio Communication
February 1993

100 WATT HF MOSFET AMPLIFIER

Build this linear for your shack

Having finally made it onto the HF bands, after years on VHF with assorted black boxes, I felt it was time to design and construct some of my own equipment. After all, a glance at my new license confirmed that amateur radio was all about "the self training of the licensee." The amplifier described here now complements a homebuilt VHF-to-20 meter transverter. It also represents a step forward in the use of low cost power MOSFETs as previously reported.[1] The amplifier uses two MOSFETs in push-pull, and is capable of producing 100 watts RF output from 1.8 to 14 MHz. Slightly lower power is available on 18 and 21 MHz.

Construction is straightforward and is certainly low cost when compared to bipolar transistors. Unlike many published designs, this unit incorporates on-board transmit/receive switching and a low-pass filter at the output. An input attenuator allows compatibility with existing equipment having an output power greater than 3 to 4 watts.

The amplifier has been designed for base station use, and requires a supply of 48 to 50 volts. While this may rule out mobile operation, it does permit the use of simple unregulated power supplies which would be impractical with 12-volt equipment. Although designed for a lower frequency limit of 1.8 MHz, testing of the prototypes has revealed the −3 dB points of the amplifier to be 225 kHz and 16 MHz representing a total bandwidth of 6 octaves!

MOSFETs or bipolars?

Before we look at construction, let us briefly consider two of the differences between bipolar

transistors and the n-channel power MOSFETs used here. Firstly, bipolar transistors are current controlled devices; i.e., a small current flowing into the base causes a much larger current to flow between the collector and emitter junctions. The FET, however, is a voltage controlled device. A voltage applied between gate and source causes current to flow between the drain and source junctions. As there is no current flowing into the gate, it can be seen by Ohm's law that the input impedance of an FET is very high. An input impedance at DC of several megohms is not uncommon. This characteristic allows very simple bias supplies to be used with FETs.

The circuit presented here (see **Figure 1**) uses a 1-k potentiometer to adjust the bias for each device. This technique cannot be used with bipolars owing to their high base current requirement. As the frequency increases, differences in input impedance become less pronounced. The FET input is highly capacitive, and this capacitive reactance causes the input impedance to be lowered at high frequencies. Also the gain of a bipolar transistor increases dramatically as the operating frequency is lowered. For example, a typical bipolar device (MRF428) has 13 dB more gain on 160 meters than on 10 meters. The average FET, however, will have a much smaller variation of gain with frequency. This allows comparatively simple networks to be used for gain compensation.

Circuit description

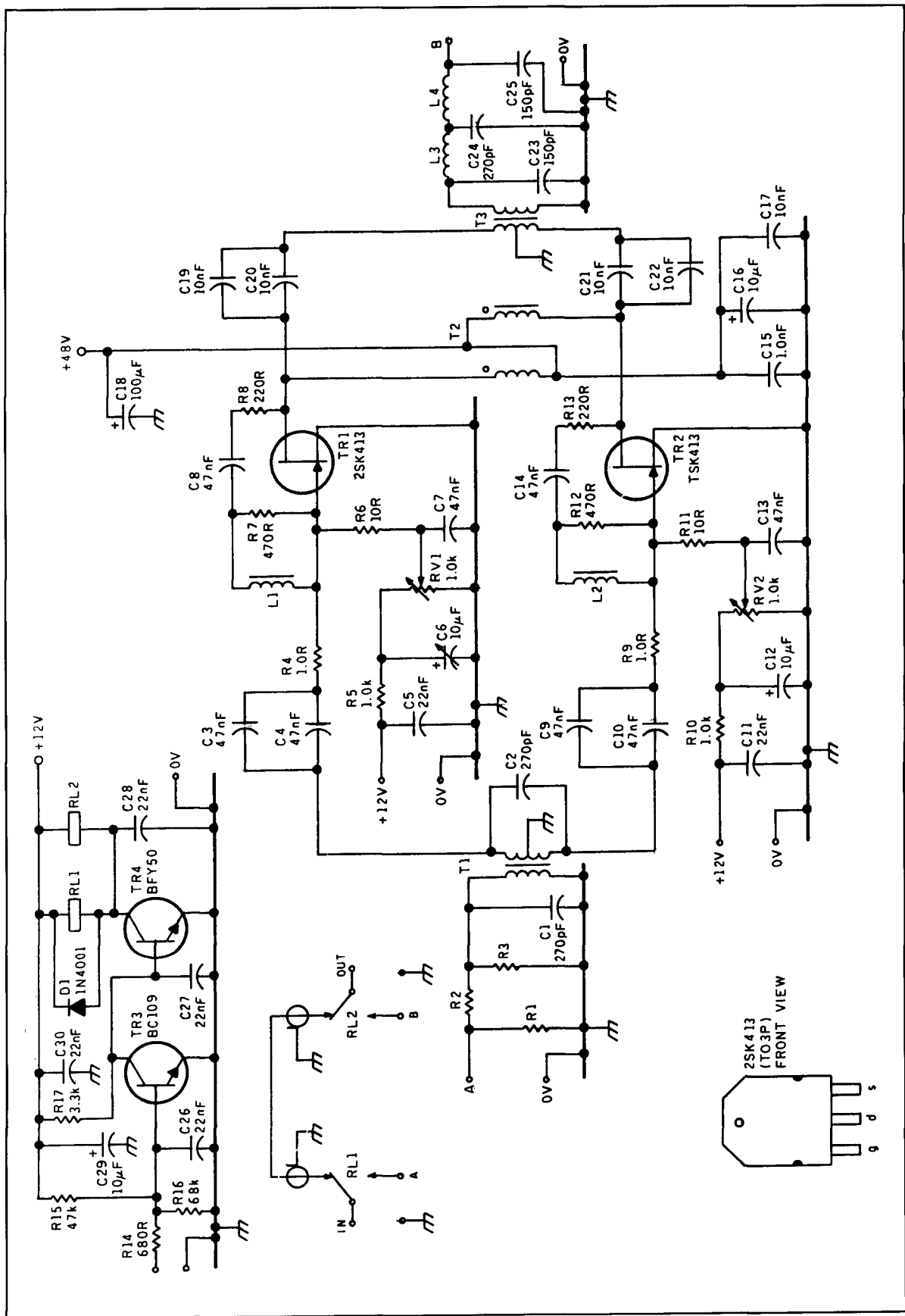The circuit has two enhancement mode power MOSFETs operating in push-pull. RF drive is routed from the input to

Figure 1. The amplifier circuit has been designed to use low-cost MOSFETs to provide a most effective way of boosting your HF signal.
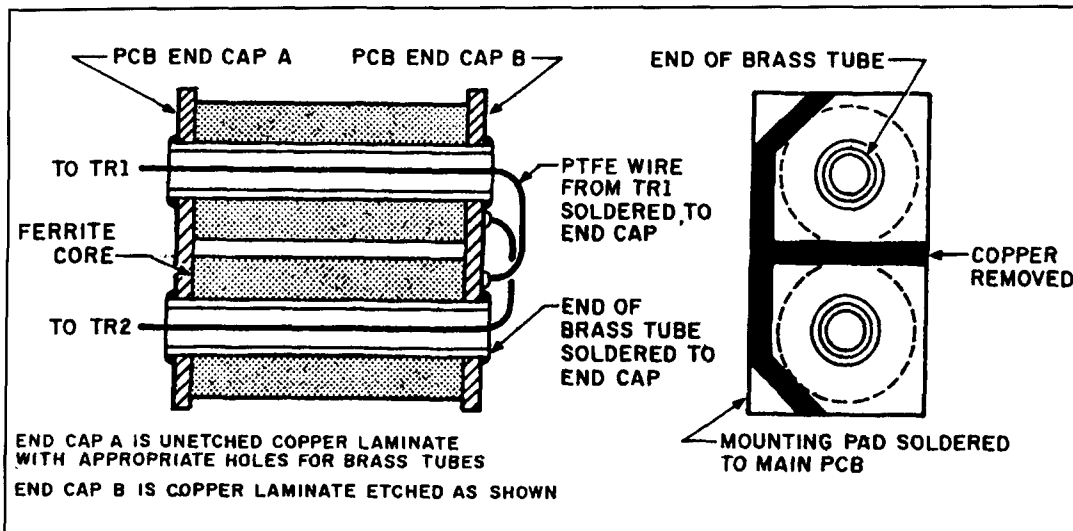
Figure 2. The output transformer R3 is constructed from two ferrite sleeves mounted on brass tubing.

transmit/receive switching relay RL1. When the amplifier is selected, the drive is applied to the 50-ohm input attenuator R1-R3. Resistor values are chosen so that the correct level of drive power is presented to the FETs. T1 is a broadband transformer (turns ratio 2:1) with two functions. Firstly, it steps down the impedance from 50 ohms on the primary (drive) side to 12.5 ohms. Secondly, the secondary winding is center tapped. This provides two antiphase signals which drive the FETs via C3, C4, and C9, C10, respectively. C1 and C2 compensate for the inductive reactance of the transformer windings and ensure a low input VSWR.

As the two halves of the amplifier are identical, I will describe the section around TR1 only. From T1, the RF passes through coupling capacitors C3 and C4. R4 prevents any tendency towards instability in TR1 by adding 1 ohm (resistive) in series with the input impedance on the FET. Bias is provided to the gate of TR1 by the combination of R5 and RV1 with decoupling by C5, C6, and C7. RV1 sets the bias voltage from 0 to 6 volts, which allows the device to be operated anywhere between class C and class A. R6 is used as a path for the bias from RV1 to the gate of TR1 and also provides a partial load for the RF input. Although this "passive-gate" technique lowers the input impedance, it allows the highly capacitive input of the FET to be driven with ease.

Negative feedback is applied to the gate by R7, R8, and L1. Coupling capacitor C8 blocks the DC voltage. The feedback from these components improves the linearity of the amplifier and provides gain compensation. At high frequencies, the feedback is primarily determined by R7, whereas at LF the reactance of L1 decreases allowing a higher level of feedback and hence lower gain. The supply voltage is

decoupled at the input by electrolytic C18 and at a point adjacent to T2 by C15, C16, and C17. T2 is a bifilar wound RF choke providing a balanced feed to the two transistors. The bifilar construction assists in the reduction of even harmonic energy.

Finally, having generated our RF signal at the drains of the FETs, we now have to combine the two halves of the signal and match to a 50-ohm output. This is achieved by broadband transformer T3, with the signal AC coupled from the drains by C19-C22. The transformer is of conventional design, with a turns ratio of 2:3 (impedance ratio of 1:2.25). Both windings use PTFE insulated wire to withstand the high temperatures that can occur in the transformer during sustained operation. Details of T3 are given in Figure 2. From T3 the output signal is passed through a five element low-pass filter, then to the transmit/receive switching relay RL2.

The Tx/Rx switching circuit is built around TR3 and TR4. The RF amplifier is switched in circuit when the input (R14) is connected to ground. This arrangement should suit most commercial and home built equipment.

Construction

The suggested layout is shown in Figure 3. Start by fitting the pins that connect the two sides of the PCB together. These should be firmly pressed into the PCB with a hot soldering iron, soldered in place on both sides and then cut close to the board. Next, assemble the output transformer T3. The transformer (Figure 2) comprises two ferrite sleeves mounted on two lengths of brass tubing. Tubes and ferrites are held in place with two small PCB end caps. Prior to soldering, the assem-
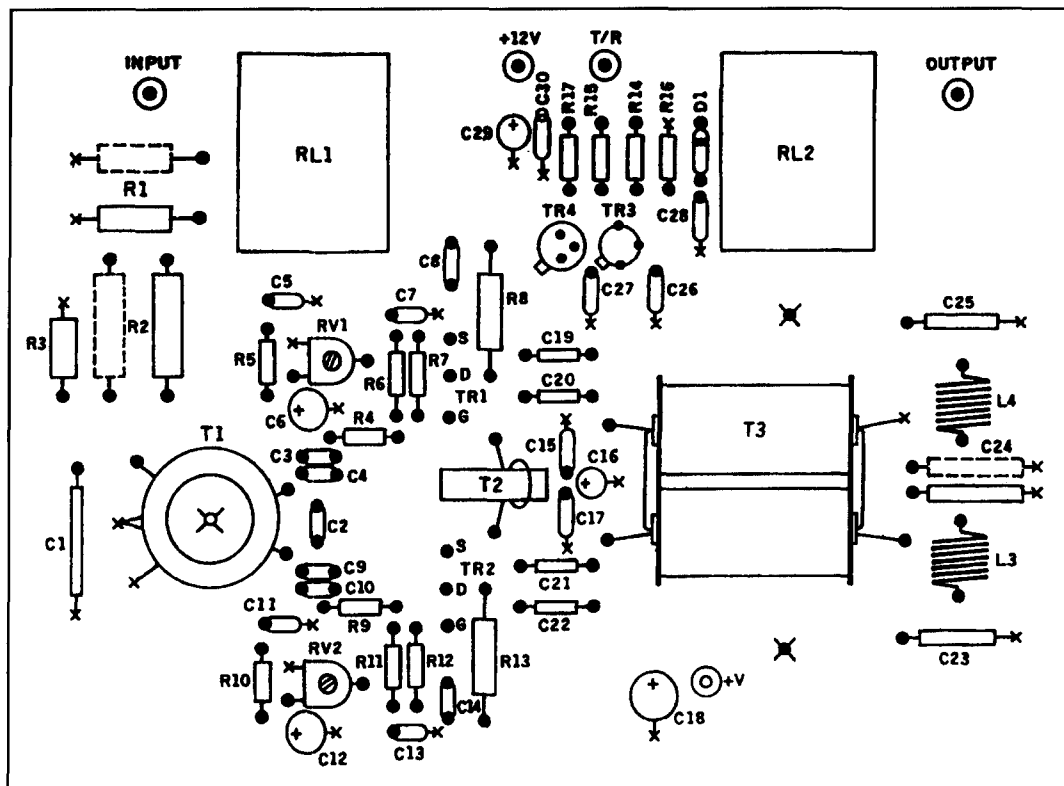
Figure 3. The prototype component layout, shown above, gives stable operation (not actual size).

Power Input (watts)	R1+R3	R2
5	820R	6R8
7.5	330R	15R
10	220R	22R
12	180R	30R
15	150R	39R

The table above is intended as a guide. Resistors should be non-inductive. They can be single units of suitable rating, or parallel combinations for increased dissipation.

Table 1. Attenuator resistor values.

bled transformer is held on a flat surface to ensure that all components are in the correct position. The tubes are then soldered around their entire circumference, and the assembled transformer (minus PTFE wire) positioned centrally between the four holes used for the termination of the windings.

With T3 placed in the correct position, solder it with a generous amount of solder. The end cap that forms the RF center tap is positioned close to C19-C22 and soldered along its entire length. Next, fit the two relays followed by the Tx/Rx switching circuit—TR3 and TR4 should be fitted within 2 mm of the PCB. The remaining PCB components can now be soldered,

leaving the remaining transformers, inductors, and the low-pass filter until last. Resistors R8 and R13 should be mounted with 1 mm clearance between the resistor body and the ground plane. The PCB also has space for a 50-ohm pi attenuator. This is used when the available input power exceeds 3 to 4 watts. If the attenuator is not required, the pads on the PCB used for R1-R3 should be bypassed with a short length of coaxial cable. The values of R1-R3 will depend on the available drive power (see **Table 1**). Further details in **References 2** and **3**.

Six connections are made between tracks on the two sides of the board. These are made in the central region of the PCB around the FET gate and drain leads. One end of resistors R6, R7 and R8 and R11, R12 and R13 are soldered on both sides of the board. If the amplifier is to be used on all bands, the low-pass filter shown in the circuit should now be fitted; this has a cutoff frequency of 22 MHz. If operation on lower frequencies is required, refer to the section dealing with testing and options. This leaves the board complete, with the exception of T1, T2, T3, L1, L2, and the FETs.

Winding the transformers

We have already looked at the construction of the core of T3. The wire recommended for
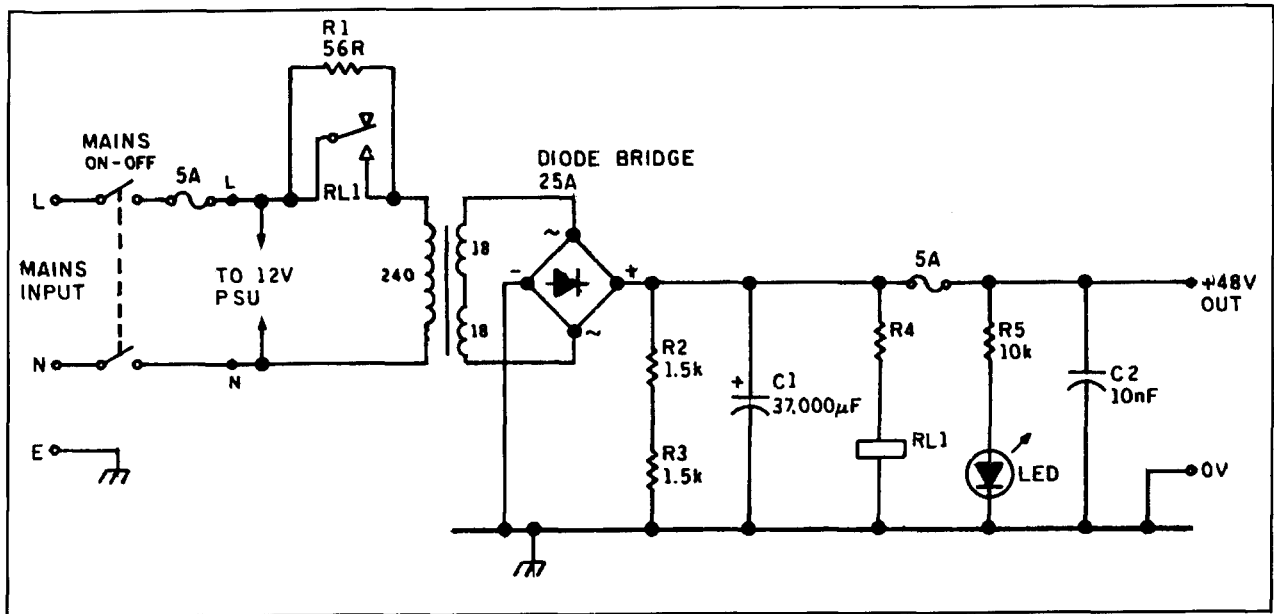
Figure 4. The resistor reduces current surges in the main power supply on power up.

the windings comprises 19 strands of 0.15 mm silver-plated copper, with an insulating sleeve of PTFE. This provides a flexible wire with high current capability; it is also resistant to corrosion and high temperatures. While this represents the ideal, any wire of similar cross section will provide identical electrical performance, but only PTFE will have the ability to withstand extreme temperatures. The circuit diagram shows the primary of T3 as a two turn, center tapped winding, capacitively coupled to the drains of the FETs. In practice, this winding consists of two half-turns of PTFE wire and one turn of brass tube.

The secondary winding is simply three turns of PTFE insulated wire would through the transformer core. One end provides the RF output to the low-pass filter, while the other is stripped over 5 to 6 mm, passed through the PCB, and soldered to ground on both sides. The balanced drain feed choke T2 is a bifilar winding on a ferrite toroid, constructed by taking a length of enamelled copper wire and folding it in the center to give two equal lengths. The center of the wire, now formed into a loop is held with a pencil. The other ends are placed in the chuck of a hand drill. The wire is twisted by turning the hand drill while pulling gently on the pencil. When the required twists have been placed in the wire, it can be removed from the drill and the best (evenly wound) section cut out for use. This twisted pair is next wound carefully on the ferrite toroid. The four ends should be cut approximately 30 mm from the ferrite and stripped of enamel over the last 20 mm.

A multimeter should be used to identify the start and finish of each winding on T2. With the circuit diagram as reference, select the two

Frequency (MHz)	2nd	3rd	4th	5th
1.8	–38	–26	–56	–42
3.5	–32	–27	–47	–48
7	–36	–23	–60	*
10.1	–32	*	*	*
14	–48	–56	*	*
18	–54	–60		
21	*	*	*	*

* = Level more than 65 dB below peak output.

Measurements made on a Marconi TRF2370 spectrum analyzer.

Table 2. Amplifier O/P harmonic content (dB).

wires that are to be joined to form the DC supply end of the choke. Be absolutely sure that the correct wires have been connected, as a mistake at this stage can prove disastrous later! Having made the choke, fit it to the board. Do not allow any point of the winding to touch the PCB's ground plane, as the high RF voltage can arc through the insulation. The construction of T1 is comparatively simple. The primary is twelve turns, and the secondary six turns with a center tap. Both are wound symmetrically over the entire circumference of the core, starting with the primary and then winding the secondary on top. The center tap is formed by winding three turns, forming the wire into a small 5 mm loop, then winding a further three turns. The loop is stripped and soldered to ground. The earthed end of the primary winding should be passed through the PCB and soldered on both sides. The inductors L1 and L2
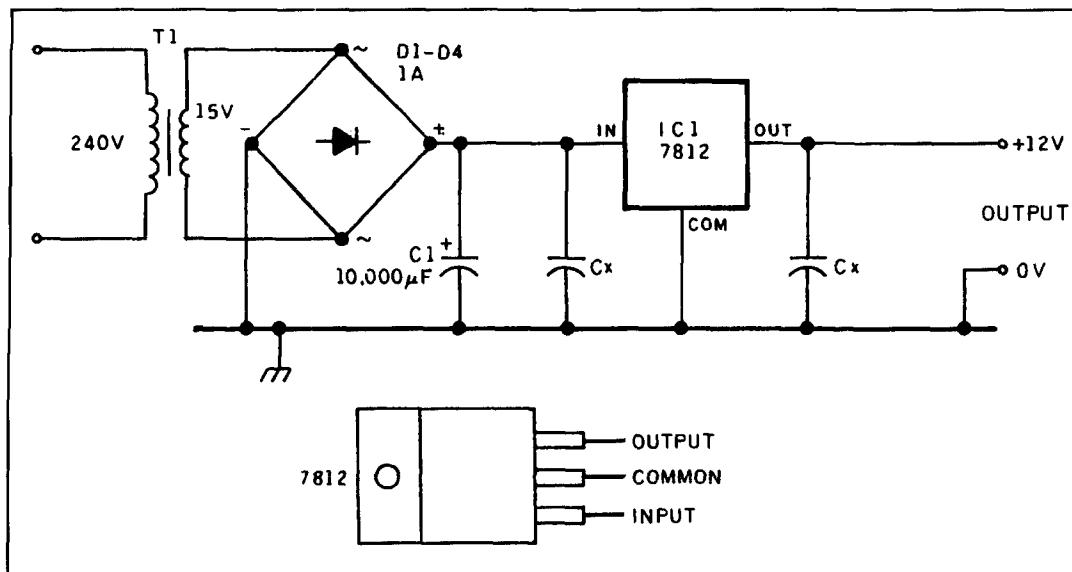
Figure 5. A simple 12-volt regulator supplies the bias and Tx/Rx switching.

can now be fitted to the reverse side of the board in parallel with R7 and R12, respectively. A short length of coaxial cable soldered between the appropriate contacts of the Tx/Rx relays provides a route for received signals, and completes the PCB.

Mounting the FETs

All that now remains is to mount the FETs on a heatsink. This is done by bending the leads of the FETs up at 90 degrees so they can pass through the appropriate holes in the PCB. Having determined the correct position for the FETs, drill the heatsink and mount the devices using mica washers and a small amount of heatsink compound. Do not tighten fully at this stage. Place the PCB onto the leads of the FETs and mark the position of the four mounting bolts that support the PCB on the heatsink. Remove the PCB and drill both this and heatsink. The PCB is held in position by 6.25 mm spacers, and after the FET mounting bolts are tightened, the leads are soldered into place. This completes the construction and the amplifier is now ready for testing.

Testing the amplifier

As the amplifier is broadband, testing is fairly simple. Firstly, the amplifier should either be fitted with the recommended 22-MHz filter, or, if this is not required, by taking the RF directly from the output of T3. A 48 to 50 volt power supply capable of 5 amps peak is required, with a regulated 12-volt 500-mA supply for bias and Tx/Rx switching. Also a 50-ohm load, SWR/power meter and a multimeter. Connect

the amplifier to the 50-ohm load via the power meter, and to the two power supplies. The 48-volt supply should have a 2-amp quick blow fuse and the multimeter in series to measure drain current.

Turn RV1 and RV2 fully *anticlockwise* (zero bias volts) and switch on the 12-volt supply. Connect the Tx/Rx pin to ground—this should cause the relays to switch to the transmit position. If successful, select the 1-amp range on the multimeter and switch on the 48-volt supply. At this stage, the ammeter should continue to read zero drain current. The only adjustment is to set the quiescent drain current to 125 mA per device. Slowly adjust RV1 until the multimeter indicates 125 mA, then adjust RV2 until the total drain current increases to 250 mA.

The amplifier can now be tested by applying approximately 1 watt of input power. On lower HF bands, this should indicate 20 to 30 watts output on the SWR/power meter. Any fault at this stage will probably be due to soldered joints or transformer winding. If all is well, change the fuse to 5 amps, adjust the ammeter range and test at full output. Run the amplifier at 100 watts for a few minutes, then disconnect the drive. Quiescent drain current will probably have increased to 300 mA. This is perfectly normal; it should, however, return to its previous value as the heatsink temperature drops.

Finally, check the input VSWR, which should be below 2:1 on all bands and usually much less. However, it can be optimized by making slight changes to the value of C1. Although this completes the testing of the amplifier, for good linearity it is important that the unit is not overdriven. The available gain at 18 and 21 MHz is less than at LF. On 18 MHz, 80 watts output can be achieved, but on 21

MHz gain is reduced to 6 to 7 dB. This represents an increase of over one S-point, and it may be possible to increase the drive power to compensate for lack of gain.

Power supply

The power supply is shown in **Figure 4** and is simply smoothed, but unregulated. RL1 and R1 give protection from current surges at switch on. With the PSU off, relay contacts are open. When 240 volts is applied to the input, current flow is limited by R1, and the charge rate on electrolytic C1 is controlled. As the voltage on C1 approaches 40 volts, the relay is energized and the resistor bypassed.

The high value electrolytic capacitor reduces the 100-Hz ripple to a low level. Linearity tests with the supply shown produced a gain compression of 0.5 dB at a power output of 112 watts. Testing with a spectrum analyzer confirmed this PSU design to be more than adequate for amateur use, and **Figure 5** uses a 7812IC to provide a 12-volt 1-amp regulated supply.

Filters and options

Harmonic output levels of the amplifier and filter are given in **Table 2**. Results compare favorably with bipolar designs and show the balancing action of the drain circuit as the second harmonic is at a significantly lower level than the third. For amateur use, all harmonics should be as low as is reasonably practicable. The five-element Chebyshev filter shown has a cut-off at 22 MHz; this provides sufficient harmonic rejection for the unit to be used directly on the higher bands.

However, below 14 MHz more filtering is needed. Suitable λ/2 filters are given in **References 2** and **3**, which reduce harmonic output to at least 60 dB below peak level. For single band operation, the on-board filter can be replaced with one appropriate for the frequency. Owing to the large physical size of air wound coils at HF, it is envisaged that powdered iron toroid inductor cores will be used, and sufficient space is available on the PCB for these. I would strongly recommend that home-brew filters are checked for low insertion loss and correct cut-off frequency before fitting.

There are many ways in which this amplifier could be developed. For example, although robust, the FETs would not survive a prolonged infinite VSWR at maximum power. The addition of a protection circuit which detected reflected power and an over temperature sensor would protect the amplifier against most fault conditions.[4]

Finally, I have found during development that the "passive gate" technique can be used with most other power FETs. As there are now a number of devices with Pd ratings of 200 watts and above, it should be possible to adapt the circuit for two or four of these FETs, increasing potential power output to the UK legal limit.

I would like to thank K. Renton, G6JMJ, at Imperial College, London, for his assistance with performance evaluation. ∎

REFERENCES
1. D. Bowman and Pat Hawker, "50 Watts RF From Low-Cost FET at 7 MHz," TT, *Radio Communication*, December 1989.
2. *Solid State Design for the Radio Amateur*, American Radio Relay League, Newington, Connecticut.
3. *The ARRL Handbook for the Radio Amateur*, American Radio Relay League, Newington, Connecticut.
4. J. Matthews, G3WZT, "Guidelines for the Design of Semiconductor VHF Power Amplifiers," *Radio Communication*, September/December 1988.

Components List

Semiconductors

TR1, TR2	2SK413
TR3	BC109
TR4	BFY50
D1	1N4001

Inductors (see text for construction details)

T1	Primary:	12 turns 0.56 mm enamelled copper wire
	Secondary:	6 turns 0.8mm enamelled copper center tapped
	Core:	Fair-Rite 59-61000401 toroid 22 mm diameter. Permeability 125

T2   Balanced drain feed RF Choke:
11 turns 0.8 mm enamelled copper, bifilar wound. 3 twists per inch. Core as T1.

T3	Primary:	2 turns comprising 5.50 mm brass tube and PTFE high temperature wire, center tapped

Components List (continued)

Secondary:	3 turns 19/0.15 mm PTFE wire	
Core:	2 x Fair-Rite 26-4354000 l ferrite sleeves (28 x 14 mm, Permeability 850)	

L1, L2	5.6 µH Toko type FL4 axial inductor
L3, L4	8 turns 12 mm ID 1 mm enamelled copper wire close wound

Resistors (All 0.25 watt carbon unless otherwise stated)

R1, R2, R3	50 ohm input attenuator
R4, R9	1R0 1 watt metal film
R5, R10	1k0
R6, R11	10R 0.5 watt
R7, R12	470R 0.5 watt
R8, R13	220R 2.0 watts
R14	680R
R15	47 k
R16	68 k
R17	3k3
RV1, RV2	1k0 min. preset

Capacitors (All ceramic unless stated)

C1, C24	270 pF silver mica
C2	270 pF
C3, C4, C7, C9	47 nF
C8, C10, C13, C14	47 nF
C5, C11	22 nF
C26, C27, C28	22 nF
C6, C12, C29	10 µF 25 volts*
C15	1 nF 1 kV
C16	10 µF 63 volts*
C17	10 nF 100 volts
C18	100 µF 63 volts*
C19, C20	10 nF 1 kV
C21, C22	0.01 µF 1 kV
C23, C25	150 pF silver mica
*Electrolytic	

Miscellaneous

RL1, RL2	12-volt single pole changeover relay (RS 344-467)	
	8 off	1 mm PCB pins
		150 mm miniature 50-ohm coax (RG 174, etc.)
		Printed circuit board
	2 off	PCB end caps for T3

The case is obtainable from H.J. Morgan Smith Ltd., Unit 3, Vernon Building, Westbourne St., High Wycombe, Bucks HP11 2PX. Telephone: 0494 532421. Price is 27.50 pounds including VAT. Carriage extra by arrangement.

Inductors and ferrite components are stocked by Cirkit (Telephone: 0992 444111).

Kits of all components (excluding case) are available from the author at Airwave Systems, 31 Benson Close, Hounslow, Middlesex TW3 3QX. Telephone: 081 472-8615.

Price for a full kit (excluding case and PSU) is 42.00 pounds. A partial kit is also available consisting of PCB, ferrites, PTFE wire, and brass tube. Price: 17.50 pounds. Note that all prices include VAT and p&p.

Thomas Cefalo, Jr., *WA1SPI*
51 Oak Street
Winchester, Massachusetts 01890

THE HAIRPIN MATCH

A review

A ntenna analysis programs have opened the door for amateurs to custom design and build their own Yagi antennas; however, glitches do occur. Designing for optimum performance, close element spacing, and short booms, or locating the antenna close to the ground, will usually result in a low input impedance. The hairpin matching network lets you transform the low impedance of such an antenna to that of the transmission line. When a proper match is achieved, VSWR is eliminated and maximum power can be transferred from the transmitter to the antenna. Redesigning the antenna for a direct match to the transmission line will, in most cases, degrade the performance of the antenna and result in decreased gain and poor front-to-back ratio.

Although the hairpin has been covered in the past,[1] I'll present a detailed review and a computer program to help you design hairpin matching networks.

Theory of operation

To use the hairpin matching network, the driven element of a Yagi antenna must be in the form of a "split element." This means the driven element is split into two halves that are insulated from each other and from the boom. In other words, the driven element is a dipole that's insulated from the antenna's boom. Some hams initially rule out using a hairpin because
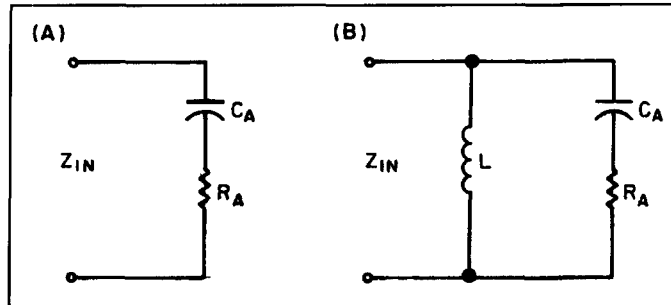
Figure 1. (A) Antenna impedance. (B) Antenna inductively loaded.

of the need for a split element. Many use a gamma match because the driven element is one piece connected directly to the boom; this is more commonly known as the "plumber's delight" construction. However, the construction of a gamma match is much more complex than that of the hairpin.

Application of the hairpin involves shortening the driven element to produce a capacitive reactive impedance. The impedance of a half-wave dipole below resonance is capacitively reactive, which is why the driven element is shortened. Changing the length of the driven element mostly affects the impedance, which becomes a resistance (RA) in series with a capacitance (CA), as shown in **Figure 1A**. The antenna is then loaded by connecting an inductor across the antenna's input terminals, as shown in **Figure 1B**. In our case, the inductor
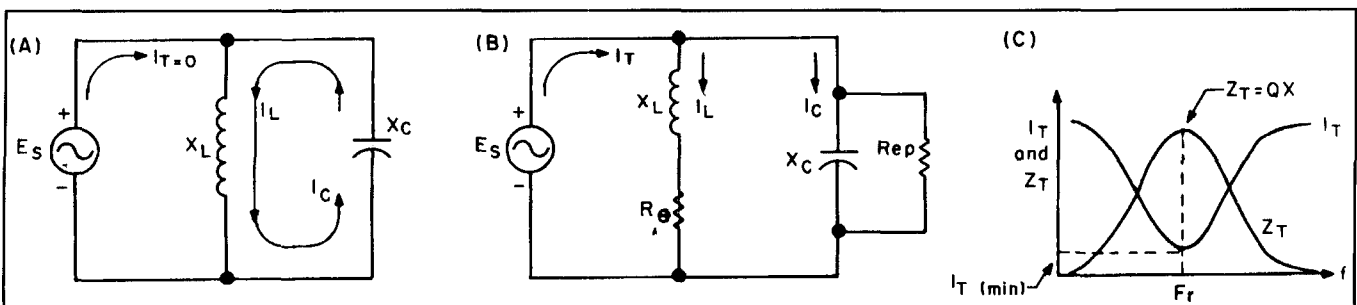
Figure 2. (A) Ideal tank circuit. (B) Practical circuit. (C) Practical circuit response.
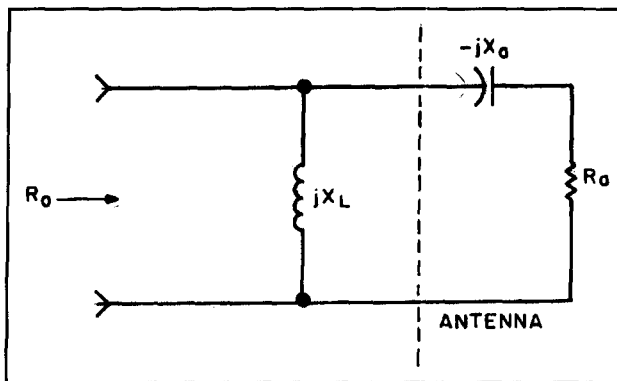
Figure 3. A matching network. If the vertical dotted line was shifted to the right of CA, the network would resemble a conventional L-matching network. (See Equations 1 and 2)

is the hairpin. It's made in the actual shape of a hairpin, hence the name.

To gain a clearer understanding of the hairpin match, let's review some basic parallel-LC circuit theory. This type of circuit configuration, also known as a "tank circuit," is shown in **Figure 2**. At resonance, the inductive and capacitive reactance are equal and opposite (180 degrees out-of-phase) and cancel each other out. Each of the branch currents I_L, I_C are also equal because $jXL=-jXC$. The total current, I_T, is zero, forcing the impedance to infinity (an open circuit) at resonance ($Z_T=E_S/I_T$ where $I_T=0$). The two branch currents circulate around the tank circuit because $I_T=0$. However, in a real circuit there is a small amount of resistance associated with each reactive element. As a result, a small current does flow at resonance. If the Q of the inductor is ≥ 10, the total current

of the parallel-LC circuit will reach a minimum value at resonance and the impedance will be some maximum value not infinity. Although the total current, I_T, is minimum, the circulating current around the tank is magnified by the Q; that is, $I_L=I_C=QI_T$. The total impedance Z_T, which is purely resistive at resonance, is Q times the reactance of either branch ($Z_T=QX$).

A parallel-resonant LC circuit is formed when the antenna is loaded with an inductor, and the antenna's resistance, RA, is in series with CA (see **Figure 1B**). In this configuration, any variation in the resistance causes the Q to change almost inversely with its value. With a small value of RA, the Q of the parallel-LC circuit will be high. At resonance, the high impedance of the parallel-LC circuit will appear to be in series with the low resistance of the antenna. The low impedance of the antenna has now been transformed to a higher level by using an inductor and taking advantage of the antenna's reactive component.

Up to this point, I've proved that a low impedance can be transformed up with a parallel-resonant LC circuit. However, a high VSWR still exists because the actual value of impedance to the match wasn't considered. To match the antenna's impedance to that of the transmission line, it's necessary to choose the proper values for L and C. The L/C ratio transforms the impedance, while the LC product determines the resonant frequency. In order to obtain a match, a balance must be struck between these two factors when their values are selected.

The following derivation can be used to determine the equations needed to calculate the values of L and C.

$$R_o = \frac{jX_L (R_a - jX_a)}{R_a + jX_L - jX_a} = \frac{jX_L (R_a - jX_a)}{R_a + j(X_L - X_a)}$$

$$R_o[R_a + j(X_L - X_a)] = X_L X_a + jX_L R_a$$

$$R_o R_a + jR_o (X_L - X_a) = X_L X_a + jX_L R_a$$

Equating Real Parts: $R_o R_a = X_L X_a$ $\quad X_L = \dfrac{R_o R_a}{X_a}$

Equating Imaginary Parts: $R_o (X_L - X_a) = X_L R_a$

$$R_o X_L - R_o X_a = X_L R_a$$

Substitute in X_L from Real Parts: $\dfrac{R_o^2 R_a}{X_a} - R_o X_a = \dfrac{R_o R_a^2}{X_a}$

$$R_o^2 R_a - R_o X_a^2 = R_s R_a^2$$

$$-R_oX_a^2 = R_oR_a^2 - R_o^2R_a$$

$$X_a^2 = R_oR_a - R_a^2$$

$$X_a = \sqrt{R_oR_a - R_a^2} \qquad (1)$$

Substitute X_a into $X_L = \dfrac{R_oR_a}{X_a}$ then

$$X_L = \frac{R_oR_a}{\sqrt{R_oR_a - R_a^2}} \qquad (2)$$

Equations 1 and **2** are the inductive and capacitive reactance required to match the antenna's impedance to the transmission line. If the vertical dotted line of **Figure 3** was shifted to the right of CA, the network would resemble a conventional L-matching network. In an L-matching network, the higher impedance is connected to the shunt-arm and the lower impedance is connected to the series-arm.

Determining the hairpin

The actual hairpin can be constructed from a heavy piece of copper wire. Equations to calculate the inductance of a hairpin aren't common or readily available. You'd need to use an inductance bridge and trial and error techniques to determine the correct dimensions for a specified inductance. Lumped-constant (coil-type) inductors can be impractical to use and may have a lower Q due to greater losses. An alternative is to use a section of transmission line as an inductor. The circuit is shown in **Figure 4**. A short-circuited section of transmission line less than a quarter wavelength will behave like an inductor. A high-Q can be obtained for a given inductance if low-loss transmission line such as open wire line is used.

The calculations that follow can be used to determine the length of transmission line required for a given inductance, and to prove that a shorted line is inductive. These equations are based on a lossless line because the hairpin

will consist of open wire line that has negligible loss. The input impedance Zin is shown below, as seen by the generator that's connected to a short-circuited transmission line.[2]

Using Trig. identity $\tan\theta = \dfrac{\sin\theta}{\cos\theta}$ then

$$Z_{in} = jZ_o \tan \beta\ell \qquad (3)$$

As **Equation 3** shows, the input impedance is a pure positive reactance. The desired value of reactance is determined by choosing the proper length of transmission line. However, unlike a lumped-constant inductor where the reactance varies linearly with frequency, the reactance of a short-circuited line varies as a function of the tangent. The reactance of a shorted line is plotted in **Figure 5**. Notice that when the length of short-circuited line is a quarter-wave long, it behaves like a parallel resonant circuit. This is why it's very important that the electrical length be much less than a quarter wavelength to maintain an inductor.

Knowing that Zin is a pure inductive reactance, we can replace Zin with XL. Using a velocity factor (V.F.) of 97.5 percent,[3] we can solve for the hairpin with **Equation 4**.

$$X_L = Z_o \tan(\beta\ell) \quad \beta = \frac{\lambda}{2\pi}$$

$$V.F. = .975$$

$$\ell(\text{meters}) = \frac{(3 \times 10^8)\ V.F.}{2\pi f} \tan^{-1}\left(\frac{X_L}{Z_o}\right) \qquad (4)$$

Now that we have an equation for the length of the hairpin, we must determine the value of Zo. Because the hairpin is based on open-wire line, **Equation 5** is the characteristic impedance of a two-wire transmission line with an air dielectric as shown in **Figure 6**.[4]

$$Z_o = 120 \cosh^{-1}\left(\frac{s}{d}\right) = 120\left[\frac{e^{s/d} + e^{-s/d}}{2}\right] \quad s \gg d \qquad (5)$$

$$Z_{in} = Z_o\left[\frac{e^{j\beta\ell} + \rho e^{-j\beta\ell}}{e^{j\beta\ell} - \rho e^{-j\beta\ell}}\right]$$

$Z_L = 0$ then voltage reflection coefficient, $\rho = -1$
B = phase constant
l = length of line

$$Z_{in} = Z_o\left[\frac{e^{j\beta\ell} - e^{-j\beta\ell}}{e^{j\beta\ell} + e^{-j\beta\ell}}\right]$$

$$e^{j\theta} = \cos\theta \pm j\sin\theta \text{ then } Z_{in} = Z_o\left[\frac{\cos\beta\ell + j\sin\beta\ell - \cos\beta\ell + j\sin\beta\ell}{\cos\beta\ell + j\sin\beta\ell + \cos\beta\ell - j\sin\beta\ell}\right] = Z_o\left[\frac{2j\sin\beta\ell}{2\cos\beta\ell}\right]$$
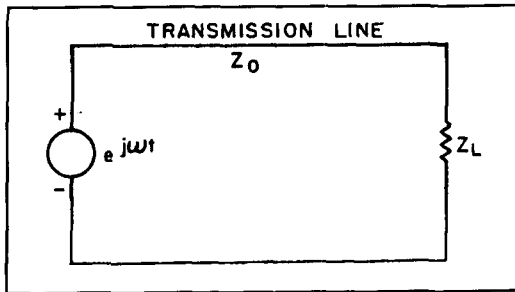
Figure 4. Using a section of transmission line as an inductor (see Equation 3).
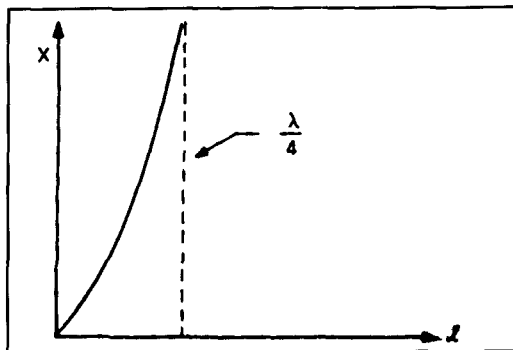
Figure 5. Reactance versus line length.
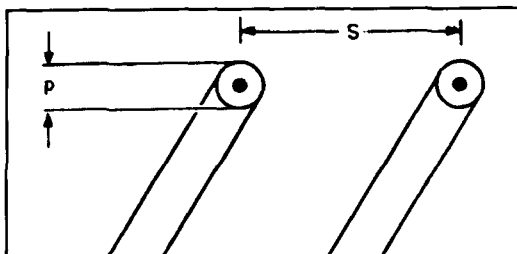
Figure 6. A two-wire transmission line with an air dielectric (see Equation 5).

```
T. Cefalo Jr.        HAIRPIN DESIGN DATA        WA1SPI

3-Element Yagi
01-12-1994   12:40:22

Frequency = 14.175 MHz
Transmission line Impedaance = 50 ohms
Antenna resistance = 22 ohms
Adjust Antenna Capacitive reactance = 24.819 ohms

Hairpin Impedance = 297.264 ohms
Equivalent Inductance = .4976 uH
Electrical Length = 8.477 degrees
Physical Length = 19.131 inches
Rod Diameter = .25 inches
Rod Spacing (center-to-center) = 1.5 inches

Impedance seen by the transmitter = 50 + j1.67E-07
VSWR = 1
```

Figure 7. Computer print-out of hairpin design data for a 3-element 20-meter Yagi.

We now have a complete set of accurate equations to construct a hairpin.

The Hairpin Match Computer Program

I wrote my hairpin match program in Microsoft QuickBasic version 4.5. It is compatible with other versions like IBM's Advanced Basic and GW Basic. A listing of the Basic program code is presented in **Appendix 1**. The program calculates the parameters needed to construct a hairpin match for a Yagi antenna based on the equations provided here. The user inputs frequency, antenna impedance (real part only), transmission line impedance, hairpin conductor diameter, and spacing. The antenna impedance can be determined with an RX bridge or an antenna analysis program. For either case, the resistance is measured with the driven element length adjusted for resonance at the design frequency.

To get an idea of how all this works, let's look at a 3-element 20-meter Yagi. **Figure 7** is a copy of the antenna's hairpin design data calculated from the program. If you want a print-out of the information, use Shift Print-Screen on your keyboard to send a copy of the computer screen to the printer.

The graph in **Figure 8** is a plot of VSWR verses frequency of the 3-element 20-meter Yagi with and without the hairpin. The impedance without the hairpin was determined using the antenna analysis program ELNEC.[5] The VSWR with the hairpin match was calculated using the antenna data from ELNEC and the equation for Ro in **Figure 3**. In this design, a broad response was desired; however, as the graph indicates, the VSWR begins to creep up toward the band edges. Although it's impossible to produce a flat VSWR within this large bandwidth, it's possible to obtain a very good compromise. If you only operate one section of a band, optimize your design for that part of the band—e.g., CW or SSB.

Construction and tuning

Using the data from the program, build the hairpin from copper or brass—preferably copper. The sliding shorting bar provides an alternative to the traditional fixed-length hairpin. This arrangement lets you vary the inductance easily. The sliding shorting bar feature isparticularly useful when you don't know the exact impedance of the Yagi, or you simply wish to fine tune the match. The complete hairpin matching network of both the fixed-length and sliding shorting bar is shown in **Figure 9**.

The end of the hairpin is connected directly to the Yagi's boom because the hairpin's center
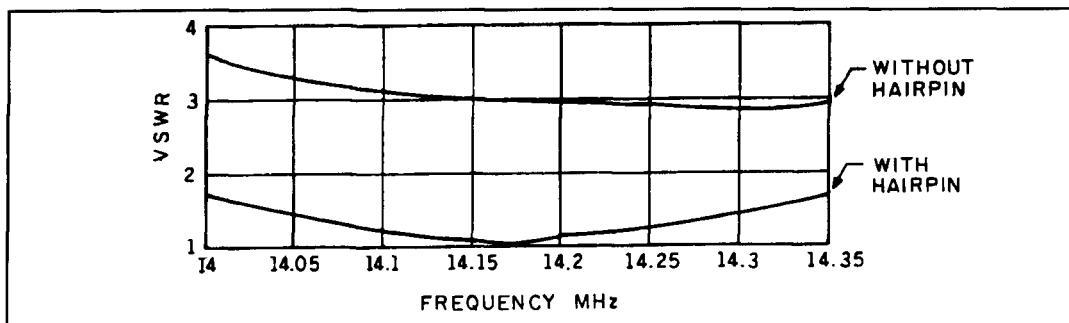
Figure 8. VSWR versus frequency for the 3-element 20-meter Yagi with and without the hairpin match.
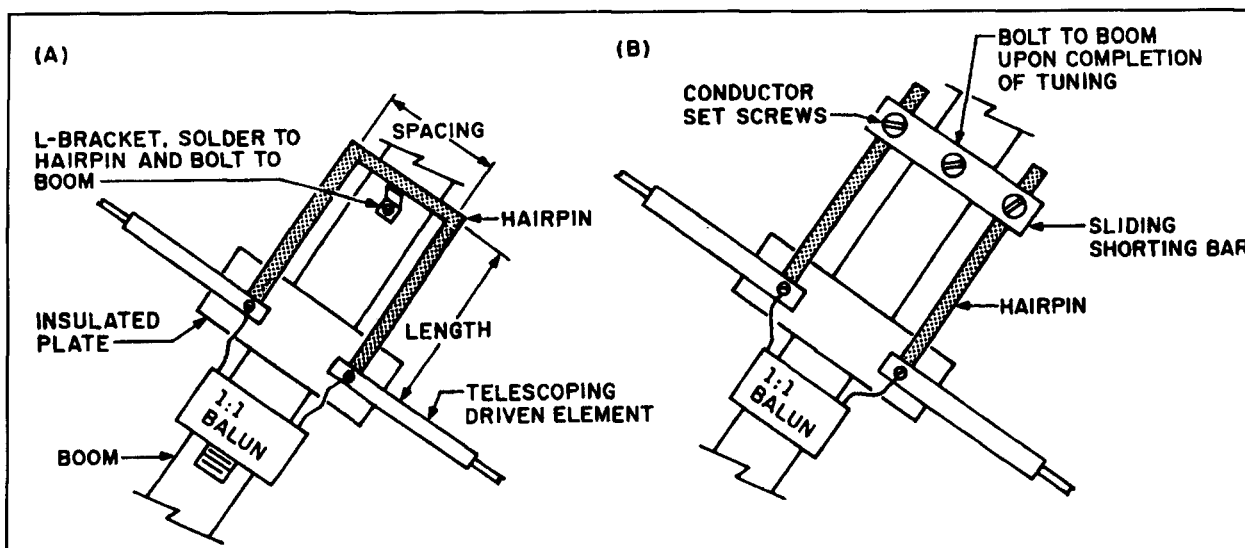
Figure 9. (A) Traditional fixed-length hairpin match. (B) Sliding shorting bar.

is an electrically neutral point. This provides a stable mechanical support for the hairpin. Also, because the center is electrically neutral, the entire antenna is kept at a ground potential that eliminates the buildup of electrostatic charges.

The feedpoint of a hairpin match is a symmetrical arrangement—a balanced feedpoint. Connecting an unbalanced coaxial cable directly to the feedpoint will result in a flow of RF currents on the outside shield of the coax. It's necessary to use a balun to choke off these currents, otherwise it will be impossible to tune or operate the antenna. You can make a 1:1 balun by sliding sleeve-type beads over the coax[3], or you can purchase one of many available commercial baluns.

The hairpin can be adjusted at ground level. Rotate the antenna so it's perpendicular to the ground; that is, pointed directly up towards the sky. Raise the antenna until the reflector is a few feet above the ground. With a VSWR bridge connected to the antenna, set the hairpin to the desired length and vary the length of the driven element, keeping each half equal in length, until the VSWR is unity. If you can't reach a match, change the length of the hairpin, and readjust the driven element again for unity

VSWR. If you use a fixed-length hairpin, it's possible to obtain a very limited range of adjustment by spreading or compressing the conductors. Spreading the conductors has the same effect as lengthening the hairpin; compressing the conductors has the opposite effect. Once you've achieved minimum VSWR, bolt the hairpin's center point to the boom. The tuning procedure is now complete and the antenna can be placed in the desired location with little or no change in the VSWR.

Conclusion

The hairpin match provides a simple but effective means of maintaining a low VSWR. It isn't difficult to build or tune a hairpin, and the process requires no special equipment. Why not consider the benefits of a hairpin matching network for your next Yagi antenna project? ∎

REFERENCES
1. J.D. Gooch, O.E. Gardner, and G.L. Roberts, "The Hairpin Match," QST, April 1962.
2. B.P. Lathi, Signals, Systems and Communication, John Wiley and Sons, 1965.
3. The ARRL Antenna Book, The American Radio Relay League, 16th edition, 1993.
4. Reference Data for Engineers, ITT, Howard W. Sams & Co., 6th edition, 1977.
5. R.W. Lewallen, ELNEC, Antenna Analysis Program, Version 3.04.

```
100 REM  ****************************************************************
110 REM * HAIRPIN DESIGN PROGRAM WRITTEN BY T. CEFALO Jr WA1SPI   @1994*
120 REM  ****************************************************************
130 CLEAR: COLOR 15,0: CLS: COLOR 14, 12
140 PRINT "T. CEFALO Jr        , HAIRPIN DESIGN PROGRAM                WA1SPI"
150 COLOR 15,0: PI=3.141592654#: PRINT
160 INPUT "ENTER FREQUENCY (MHz)";F: F=F*1000000!
170 INPUT "ENTER ANTENNA FEED-POINT RESISTANCE (ohms)";RA
180 INPUT "ENTER TRANSMISSION LINE IMPEDANCE (ohms)";ZO
190 IF RA>=ZO THEN GOTO 480
200 INPUT "ENTER HAIRPIN ROD DIAMETER (inches)";D
210 INPUT "ENTER HAIRPIN ROD SPACING (inches)";S
220 PRINT: INPUT "ENTER TITLE OF DESIGN (return for none)";TITLE$
230 XA=SQR((ZO*RA)-RA^2): XL=ZO*RA/XA: A=S/D
240 ZHP=120*LOG(A+SQR(A*A-1)): EL=ATN(XL/ZHP)*180/PI
250 PHYL=(3E+08*EL*39.37*.975)/(F*360)
260 ZIN=SQR((XA*XL)^2+(XL*RA)^2)/SQR(RA^2+(XL-XA)^2)
270 THETA=ATN(RA/XA)-ATN((XL-XA)/RA): L=XL/(2*PI*F)
280 RE=ZIN*COS(THETA): IM=ZIN*SIN(THETA)
290 K=RE/ZO: P=SQR((K-1)^2+IM^2)/SQR((K+1)^2+IM^ 2)
300 V=(1 + P)/(1- P)
310 CLS : COLOR 11,12: D$=DATE$: T$=TIME$
320 PRINT "T. CEFALO Jr           HAIRPIN DESIGN DATA                WA1SPI"
330 COLOR 15,0: PRINT : PRINT TITLE$
340 PRINT D$,T$: PRINT
350 PRINT "FREQUENCY="; F/1000000!;"MHz"
360 PRINT "TRANSMISSION LINE IMPEDANCE=";ZO;"ohms": PRINT
370 PRINT "ANTENNA RESISTANCE=";RA;"ohms"
380 PRINT "ADJUST ANTENNA CAPACITIVE REACTANCE=";XA;"ohms": PRINT
390 PRINT "HAIRPIN IMPEDANCE=";ZHP;"ohms"
400 PRINT "EQUIVALENT INDUCTANCE=";L/.000001;"uH"
410 PRINT "ELECTRICAL LENGTH=";EL;"degrees"
420 PRINT "PHYSICAL LENGTH=";PHYL;"inches"
430 PRINT "ROD DIAMETER=";D;"inches"
440 PRINT "ROD SPACING (center-to-center)=";S;"inches": PRINT
450 IF SGN(IM)=-1 THEN A$="-j" ELSE A$="+j"
460 PRINT "IMPEDANCE SEEN BY THE TRANSMITTER=";RE;A$;ABS(IM);"ohms"
470 PRINT "VSWR=";V:END
480 PRINT: PRINT "Antenna impedance must be less than transmission line"
490 INPUT "EXIT PROGRAM (Y/N)";D$: IF D$="N" OR D$="n" THEN GOTO 130
500 CLS: END
```

PRODUCT INFORMATION

Programmable Low-Pass Bessel Filter IC Has 5Hz To 13MHz Cutoff

Analog Devices' AD896 fully programmable filter IC features a 7th-order Bessel filter with programmable cutoff frequency (5 Hz to 13 MHz) and equalization/boost amplitude (0 to +9 dB). These parameters are set using individual analog control lines. The filter provides group delay characteristics, with flatness of ±750 ps over the complete operating range. Applications for this 16-pin device include disk-drive read channels, communications receivers, and data acquisition systems. A factory "metal-mask" option allows the filter to be changed to equiripple characteristics. This IC is a pin-for-pin compatible replacement for the SSI 32F811.

The AD896 features a fully differential signal path with both differentiated and normal (single-ended) low-pass outputs. These outputs have matched time delays to ensure time coherence of the outputs. Cutoff frequency and equalization are trimmed to an accuracy of ±10 percent.

The AD896 operates from a single +5V supply and is available in plastic DIP and SOIC packages. Typical power consumption is 0.3W; in power-down mode (set by a control line) consumption drops to 50 mW. Prices begin at $3.95 (1000s).

For more information contact Analog Devices, Inc., 181 Ballardvale Street, Wilmington, MA 01887, or call 617-937-1428; Fax: 617-821-4273.

Peter O. Taylor
AAVSO Solar Division
4523 Thurston Lane, #5
Madison, Wisconsin 53711-4738

THE SOLAR SPECTRUM

Solar Flares... Who Cares?

In the interests of self-preservation, we're going to reply to the question that leads off this column before delving into the topic itself. The answer, among others, is: astronauts and future space travelers, solar astronomers, and scientists in space-related fields; make no mistake about that!

The reasons are familiar to most of us. The X-rays produced in many solar flares and spewed throughout interplanetary space can spawn short-lived effects at the Earth, and raise legitimate concerns for anyone who ventures beyond the protection of the atmosphere. And, the energetic processes and peripheral effects of flares continue to be focal points of interest for solar and atmospheric physicists, and for others who study the Sun. Now however, a new idea championed by Dr. John T. Gosling,[1] a scientist at Los Alamos National Laboratory, is stirring increased attention among those who examine the source and affect of solar activity.

Near the maximum of each solar cycle, great storms plague the Earth's magnetic field and dazzling aurorae can be viewed from much
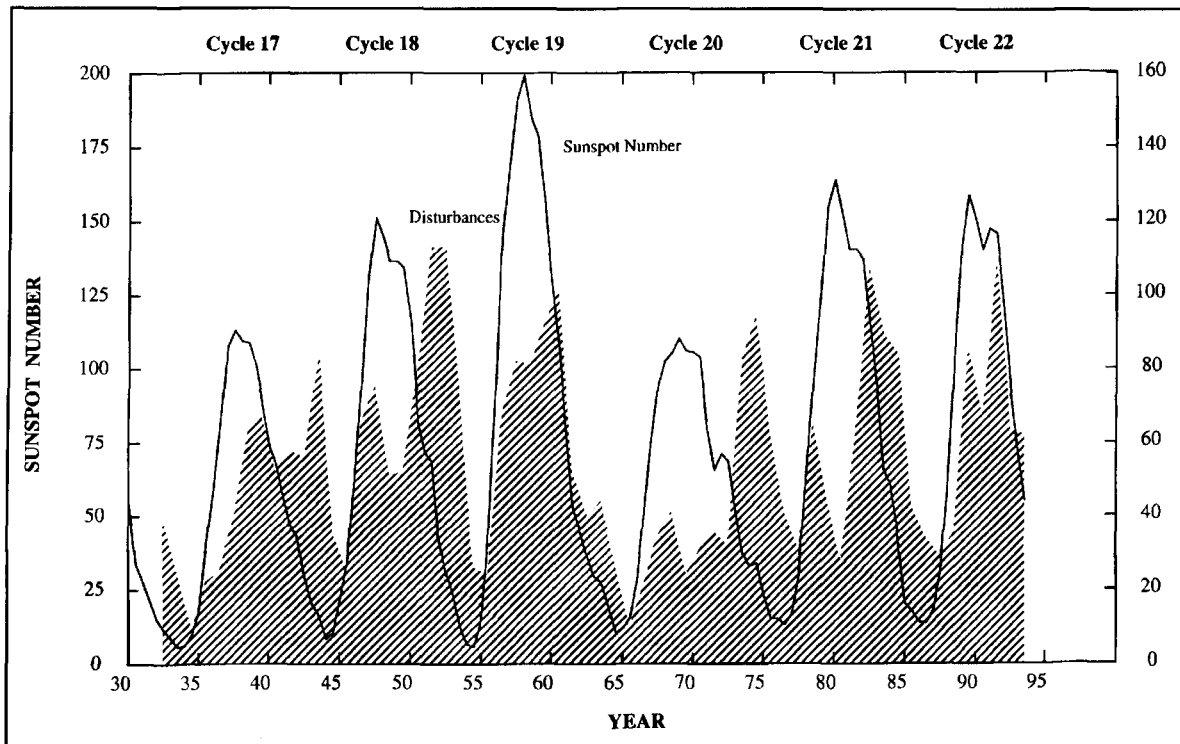
Figure 1. The variation of the yearly averaged relative sunspot number since 1930 (solid line), and the variation in the number of geomagnetically disturbed days per year when the planetary magnetic disturbance index (Ap) was greater than or equal to 25, since 1932 (shaded area). Note that the number of such storms typically peaks twice: first, around sunspot maximum (there is some tendency for the amplitude of this peak to correlate with the amplitude of the solar cycle), and a second peak several years after sunspot maximum—the so-called "precursor" peak. Diagram and comments courtesy of R.J. Thompson, IPS Radio & Space Services, Sydney, Australia.
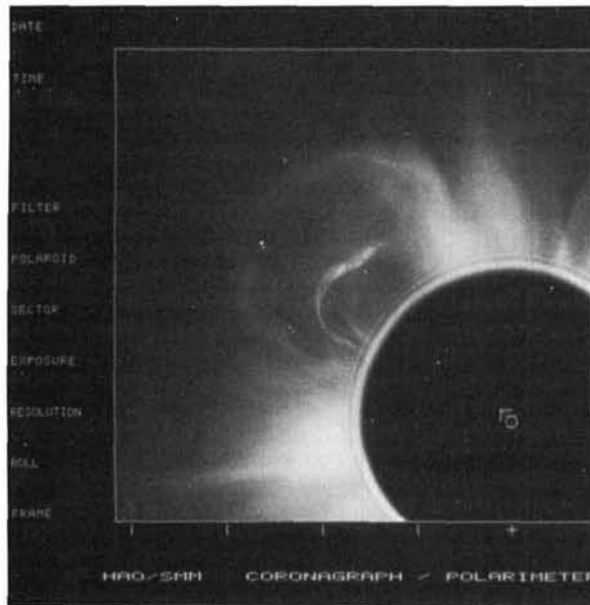
Photo A. A coronal mass ejection lifting off the Sun. The outer loop is the leading edge of the CME, while the bright inner loop is an erupting prominence. Photograph courtesy of A.J. Hundhausen, National Center for Atmospheric Research.
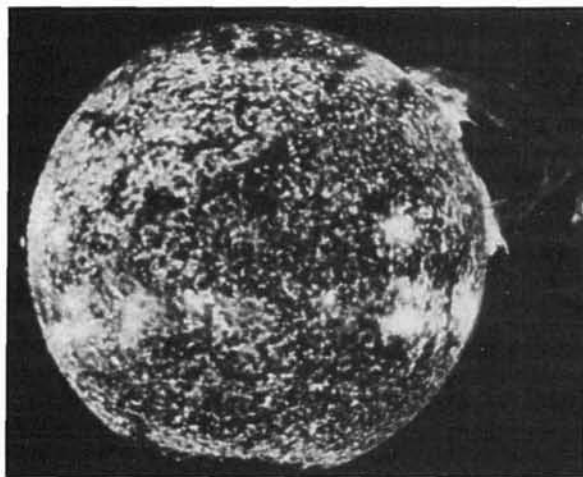
Photo B. Computerized color enhancement of an erupting prominence taken by the solar telescope on board the Skylab spacecraft. Photo courtesy of NASA.

lower latitudes than normally possible. Two important events during the mid-nineteenth century significantly influenced scientific thinking concerning the source of such storms: First, the number of magnetic disturbances was linked to the newly discovered sunspot cycle (e.g., **Figure 1**), and second, the English astronomer Richard Carrington observed the first solar flare and suggested that it might have some association with a spectacular geomagnetic storm that followed.

Shortly thereafter, evidence began to accumulate that seemed to confirm flares as the

source of terrestrial magnetic disturbances. Since then the two phenomena have been reasonably well correlated, and the relationship has been generally accepted for more than half a century. In spite of such circumstantial evidence, however, the flare/magnetic storm connection has never been observed on a one-for-one basis. Some powerful flares—even those which are centered in the western solar hemisphere where the interplanetary magnetic field should make connection with the Earth most likely—do not result in disturbances, and some storms do not appear to be associated with a particular flare.

The long-standing link between flares and geomagnetic storms is a bit like the story of the boy whose mother sent him to a psychiatrist because he was continually snapping his fingers. "Why do you snap your fingers?" asked the doctor. "To keep the snakes away!" countered the boy. "But there are no snakes..." argued the doctor. To which the boy replied, "That's because I keep snapping my fingers!"

The flare-origin of large geomagnetic storms began to be called into question over fifteen years ago, when many disturbances were tracked back to the Sun and found to originate with the solar wind and coronal holes. (We discussed these phenomena in the Summer 1992 issue of *Communications Quarterly*.) During the ensuing years, storms that repeat at about 27-day intervals (the apparent rotational period of the Sun) were unequivocally tied to these features. Such recurrent geomagnetic disturbances take place as matter streaming into space from a low-density portion of the Sun's atmosphere (i.e., from a coronal hole), co-rotating with the Sun, returns to a geoeffective position.

However, until recently the occurrence of large *non-recurrent* storms—those that are often accompanied by the spectacular terrestrial effects we read about and hear described by the popular media—has continued to be linked to the energy bursts from especially violent solar flares, or occasionally to transient coronal holes.

Gosling has now offered convincing evidence for an alternative theory; one which relegates the flares that seem to precede many of these storms to a secondary effect triggered by gigantic ejections of matter from the Sun's atmosphere, and identifies the latter as the true source of all great geomagnetic disturbances. Moreover, Gosling's research cites the strength, speed, and configuration of the magnetic fields carried within the ejected material—and *not* simply energetic solar particles—as the causal agent of all such magnetic storms.

As with the solar wind, the release of immense quantities of coronal material into space (up to eleven thousand million tons of solar matter in some coronal ejections) are rela-

tive latecomers to solar physics. The expulsions—known as coronal mass ejections, or simply "CMEs"—are primarily detectable from spacecraft-mounted coronagraphs that employ sensitive electronic detectors in place of photographic film to record them.

CMEs were first observed in 1973 by instruments aboard the U.S. solar satellite OSO-7, and have subsequently been explored by a number of spacecraft including Skylab, the Naval Research Laboratory P78-1, and Solar Maximum Mission satellites. These observations have been complimented by others in the inner corona from ground-based instruments at Mauna Loa Solar Observatory in Hawaii, and by the zodiacal-light photometers on board the Helios spacecraft.

Although coronal mass ejections exhibit a number of morphological features, they often have the appearance of giant loops or bubbles of solar material rooted to the photosphere (**Photo A**). A more technical definition was provided by National Center for Atmospheric Research astronomer A.J. Hundhausen in 1984: ". . .an observable change in coronal structure that (1) occurs on a time scale between a few minutes and several hours and (2) involves the appearance of a new, discrete, bright, white-light feature in the coronagraph field of view."[2] An extension of this definition includes the requirement that the event display a predominantly outward motion.

In any event, CMEs are mammoth structures that can occupy up to a quarter of the solar limb before lifting off into space, frequently accompanied by the remnants of an erupting prominence (**Photo B**). They appear to originate mainly in closed magnetic field regions within the coronal streamer belt (**Photo C**) and are thought to arise from changes in the large-scale coronal magnetic field.

Flares on the other hand, result more from the reorganization of small, but intense fields generally located near active sunspot regions. It is interesting that CMEs can occur in both normal sunspot/flare zones and at much higher solar latitudes. Moreover, their occurrence seems to follow the traditional butterfly-pattern; appearance at low heliographic latitudes during cycle minimum ranging to high latitudes around maximum (**Figure 2**).

It has long been clear that CMEs are unlikely to be a flare-induced phenomenon. Solar flares accompany a fair portion of these mass ejections, but interestingly, those CMEs usually begin their departure *before* the flare occurs; a factor that has contributed to some astronomers' belief that CMEs—and not flares— are the crucial link between the solar disturbance, its propagation through the heliosphere, and effects at the Earth. Further evidence

Photo C. This striking view of the solar corona was prepared from data acquired by NASA's Solar Maximum Mission Satellite. The prominent feature extending from the Sun is a coronal streamer; several others are also visible. The large streamer extends outward for more than 1,000,000 miles. The white dots (camera imperfections) are the approximate size of the Earth. Photograph courtesy of NASA.

of their disassociation lies in the observation that flares which occur with CMEs typically erupt far to one side of the broad-based mass ejection. Of course, solar flares occur in much larger numbers than do CMEs, so only a portion of flares are initiated by these events.

The way that those particular flares are caused is a matter of conjecture. It has been suggested that when some CMEs lift-off, segments of the massive bubble—containing oppositely directed magnetic fields—become elongated and temporarily remain attached to the Sun. As the fields are distorted and brought together during the ejection's exodus, they may reconnect and cause the flare. However, the exact mechanism remains unclear at this juncture.

The leading edges of fast-moving CMEs— about one-third of all these events—drive giant shock-waves before them, through the solar wind at speeds up to 1200 kilometers per second. (Since they typically move at or below the speed of the solar wind, low-speed CMEs do not drive shock disturbances.) Although flares certainly impart energy to solar particles, CME-initiated shock-waves are capable of energizing particles on a much more impressive scale.

As the trailing mass of plasma moves out into interplanetary space, it can grow to an enormous size. Eventually, some of these assemblages expand to encompass a greater amount of space than the Sun itself; far larger than the narrow cone covered by outward-moving flare particles. When the shock-wave and cloud of material reaches the Earth and conditions are right, the geomagnetic field can undergo a major disturbance.
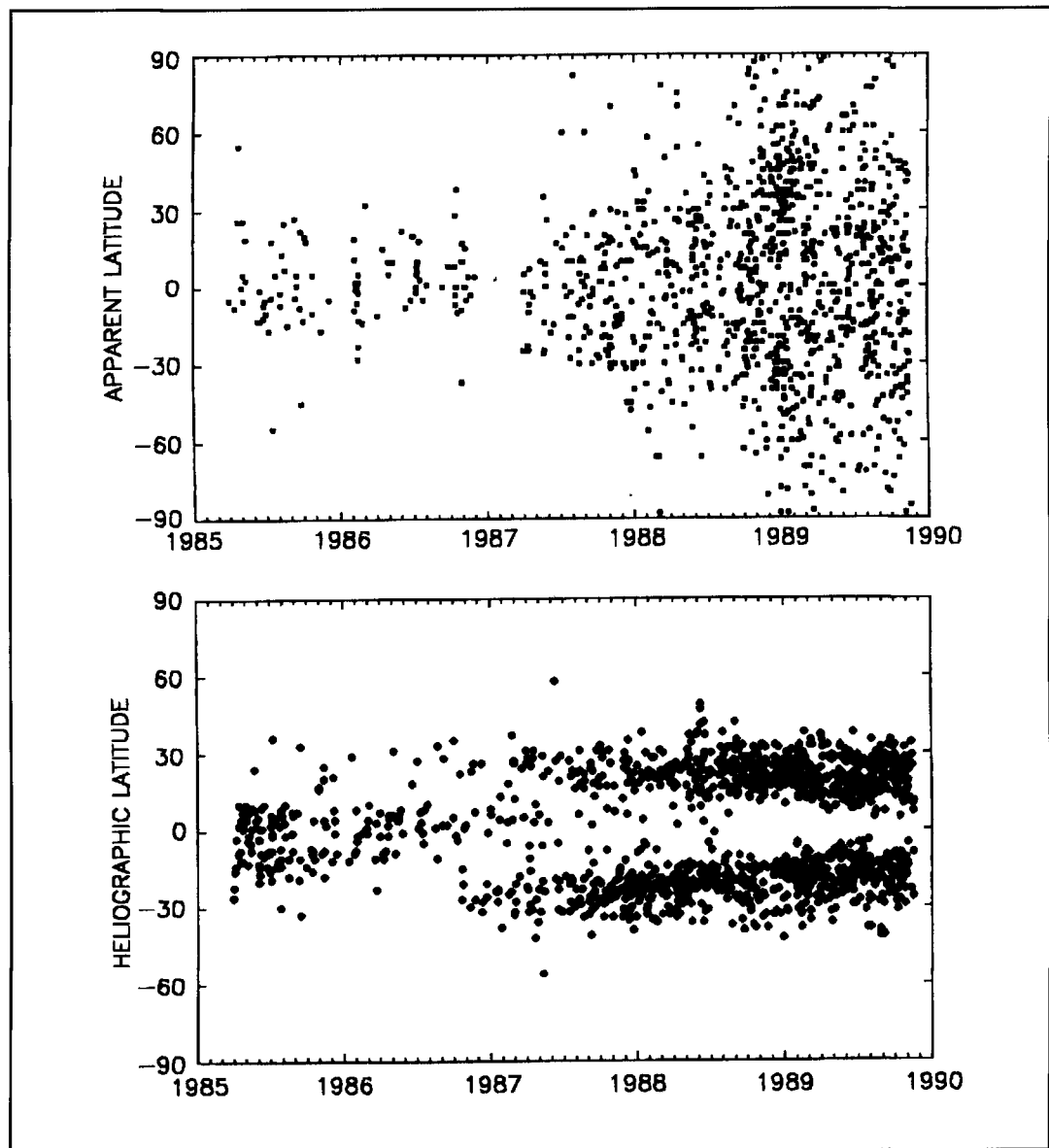
Figure 2. Butterfly-diagrams showing the locations of coronal mass ejections[3] (top) and emerging sunspot groups between April 1985 and November 1989. Note that while both phenomena follow the traditional butterfly pattern—appearance at locations near the solar equator at cycle minimum (September 1986) progressing to high latitudes at maximum (July 1989)—CMEs also occur at high latitudes. CME locations are "apparent" latitudes because they are computed from position angles along the Sun's limb.

As with most of the Sun's active phenomena, the number of CMEs appears to rise and fall with the solar cycle. Gosling reports that, on average, the Earth intercepts about six CMEs per month at cycle maximum, and less than one for any month around minimum. When the origins of the thirty-seven major and large geomagnetic storms that occurred during the maximum of cycle 21 are analyzed, it appears likely that all but one were associated with the passage of a shock, a CME, or both.

The direction of the looping interplanetary magnetic fields driven by fast CMEs plays an important role in the new scenario. If the fields

happen to point north as the shock and plasma cloud pass the Earth, nothing much happens. The geomagnetic field simply deflects the flow around the magnetosphere. However, when the fields are directed southward (opposite to the Earth's field), they reconnect with the terrestrial field, and the possibility of magnetic storms and aurorae rises in accordance with the speed and strength of the rapidly moving flow.

Some very complicated magnetic and electrical processes are then set into motion within and around the magnetosphere. The energy spawned by this mechanism drives the ensuing magnetic storm and accelerates both solar and

TRADITIONAL

rapidly evolving solar magnetic field
↓
reconnection (?)
↓
————————————— FLARE —————————————

↓	↓
rapid heating	particle acceleration
↓	↓
thermally driven MATERIAL EJECTION	solar storage and propagation to Earth
↓	↓
INTERPLANETARY SHOCK DISTURBANCE	SOLAR PROTON EVENT
↓	↓
GEOMAGNETIC STORM ↓ AURORA	POLAR CAP. ABSORPTION EVENT

MODERN

evolving solar magnetic fields —————
↓   ↓
instability (?) buoyancy (?)   reconnection (?)
↓   ↓
——— CORONAL MASS EJECTION ———————   ↓
↓   ↓   ↓   ↓   ↓
low speed high speed ERUPTIVE PROMINENCE reconnection (?) ↓
↓   ↓   ↓
INTERPLANETARY SHOCK   FLARE

low speed / high speed		reconnection (?)	
↓	↓ particle acceleration	↓ particle acceleration	
↓	↓	↓	
CME directed ~earthward	CME from visible hemisphere	flare in western hemisphere	
↓	↓	↓	
GEOMAGNETIC STORM & AURORA	GRADUAL PROTON EVENT	IMPULSIVE PARTICLE EVENT	
↓	↓	↓	
delay: 1-4 days	minutes-day	minutes	
↓	↓	↓	
duration: days	days	hours	

Figure 3. Traditional and modern paradigms of the chain of events leading to geomagnetic storms and other phenomena. Capital letters depict observational occurrences and lowercase letters indicate physical processes or descriptive characteristics. From diagrams that originally appeared in Gosling (1993).[1]

local particles down through the magnetosphere and into the Earth's upper atmosphere, where they strike atoms and molecules. As we know, when struck, the atmospheric components become "excited" (ionized—raised to a higher energy state) and begin to glow, resulting in the aurorae (**Photo D**).

Unfortunately, many geomagnetic storm forecasters (**Photo E**) are not yet privy to regu-

lar CME data. Therefore, they must look for more visible signs of magnetic instabilities on the Sun (such as disappearing filaments) and base their predictions on the observation of flares and other events that may actually be secondary to the true storm source. Adding to the difficulty, those CMEs that would affect the terrestrial environment most dramatically— events which erupt while facing towards the

Photo D. Aurorae photographed from the continent of Antarctica. Photo courtesy of M. Taylor, University of Wisconsin.

Photo E. Geomagnetic storm forecasters at the IPS Radio & Space Service in Sydney, Australia. This warning center is one of several strategically located centers scattered throughout the world. Photograph courtesy of R.J. Thompson.

terrestrial neighborhood—are difficult to measure from locations near the Earth. In fact, Gosling suggests that at some future time CMEs be monitored by spacecraft stationed at the L4 and L5 Langrangian points, where the view will be unobstructed by the Sun's glare.

Figure 3 shows the series of events that are kindled by magnetic deformations at the Sun, according to both traditional and modern scenarios. A new solar-terrestrial exhibit planned for the *Smithsonian Air and Space Museum*, and titled, "Electric Space: Our Sun-Earth Environment," will emphasize the newly recognized importance of coronal mass ejections in our understanding of the Sun's influence on the terrestrial environment. ∎

REFERENCES

1. J.T. Gosling, "The Solar Flare Myth," *Journal of Geophysical Research*, 98, Number A11, pages 18, 937–949, 1993.

2. A.J. Hundhausen, et al., "Coronal Mass Ejections Observed During the Solar Maximum Mission," *Journal of Geophysical Research*, 89, Number A5, 2639–646, 1984.

3. J.T. Burkepile and O.C. St. Cyr, "A Revised and Expanded Catalogue of Mass Ejections Observed by the Solar Maximum Mission Coronagraph," *TN-369+STR*, National Center for Atmospheric Research, 1993.

Dick Weber, PE, *K5IU*
2 Hickory Creek Drive
P.O. Box 44
Prosper, Texas 75078

AERODYNAMIC BALANCING: PART 1

Yagi antennas with slightly drooped horizontal elements

A t this time, there are several schemes being used to minimize the torque required to turn Yagi antennas subject- ed to a wind stream. The goal of these methods is to "aerodynamically balance" the antenna so there are no angles relative to the wind where the rotor stalls or has a hard time turning the antenna. The ultimate aim is to turn antennas with ease, regardless of wind speed. If a perfect balance, or very near perfect balance, were to be achieved, small rotors could be used to turn large arrays. Although several individuals have claimed success in reducing torque imbalances, a thorough, basic, analysis is needed in view of recently published information.[1] The article, "Determination of Yagi Wind Loads Using the Cross Flow Principle," shows that the method currently used to determine Yagi antenna wind loads is incorrect. Because current balancing schemes are based on an erroneous method of determining wind loads, it is apparent that a

balancing technique using correct wind loads is required. Such a technique is presented here.

Yagis with highly drooped elements, like 7 and 3.5-MHz Yagis with full-size elements, have different primary loading effects than Yagis with slightly drooped elements. As a result, other means must be used to obtain a balanced, or less imbalanced, condition. There is no straightforward balancing methodology that applies to Yagis of all sizes. This article is aimed at Yagis with slightly drooped horizontal elements typically found in Yagis for 10 MHz and higher, and Yagis with shortened elements for 7 MHz.

Wind loading

The faulty "variable area" wind loading method now in current use assumes that the force developed on a tube in a wind stream is
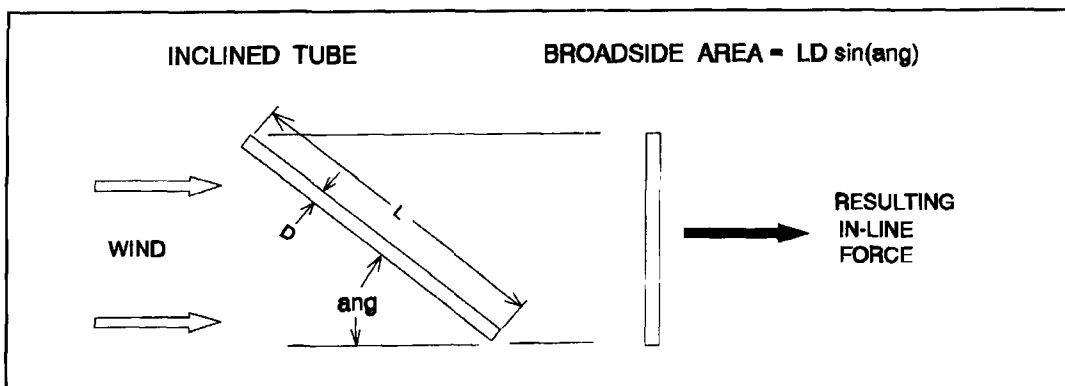
Figure 1. Assumed force direction based on "variable area" method.

INCLINED TUBE

RESULTING FORCE
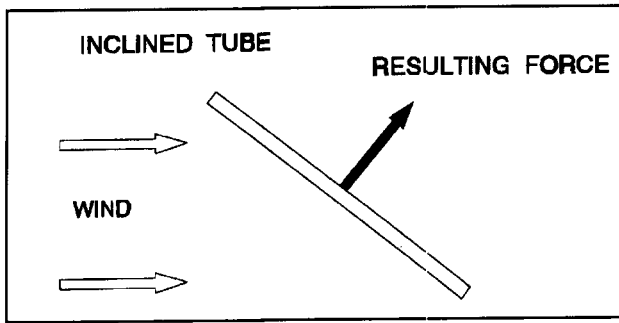
WIND

Figure 2. Correct force direction based on "cross flow principle."

produced only in the direction of the wind. The magnitude of the resulting force is a function of the tube's area broadside to the wind stream as shown in **Figure 1**.[2,3] **Equation 1** is used with the "variable area" method to find the wind force magnitude. **Reference 1** showed that the resulting force is actually perpendicular to the axis of the tube as illustrated in **Figure 2**. The fact that the resulting wind force on a tube, or any other shape, is perpendicular to the tube's axis, drastically alters current balancing methods. Further, the article develops an equation to determine the magnitude of the wind force based on the "cross flow principle," which is shown in **Equation 2**.[4,5] It is important to point out that the resulting wind force is always perpendicular to the tube's axis, and the magnitude of the force varies as a function of the angle of attack described by **Equation 2**. This force is the result of a distributed wind load over the length of the tube with the net effect realized at the tube's midpoint.

$$Fp = C A P \sin (ang) \qquad (1)$$

Fp = wind force in line with wind(by "variable area" method)
C = tube drag coefficient

P = dynamic wind pressure
P = 0.00256 V^2 (psf)
V = exact wind speed (MPH)
A = tube projected area (sqft)
A = length x diameter = L D
L = length (ft)
D = diameter (ft)
ang = angle of attack (deg)

$$Fcf = C A P \sin^2(ang) \qquad (2)$$

Fcf = wind force perpendicular to tube axis (by Cross Flow Principle)
C = tube drag coefficient
P = dynamic wind pressure
P = 0.00256 V^2 (psf)
V = exact wind speed (MPH)
A = tube projected area (sqft)
ang = angle of attack (deg)

Now that wind force magnitude and direction for a tube are established, a procedure to wind balance Yagi antennas can be developed.

The concept of torque

There are two very different physical effects that cause things to twist or turn. Although both have the same units of measurement, they are quite different and are not interchangeable concepts. These concepts are torque and moment.

Torque is the twisting or rotating potential transmitted down the length of a shaft or member when that shaft or member is rotated about its axis, or when rotation is attempted. In an automobile engine, torque is one of the engine rating factors that indicates how "hard" the crankshaft can twist or how much rotational potential can be produced.

Moment describes the bending effect of a beam or other structural member due to an applied force, its own weight, moments trans-

MOMENT

W

WEIGHT CREATES MOMENT
IN STRUCTURAL MEMBER

MOTOR

SHAFT

FAN

SHAFT

TORQUE

TWISTING LOAD ON SHAFT
CREATES TORQUE
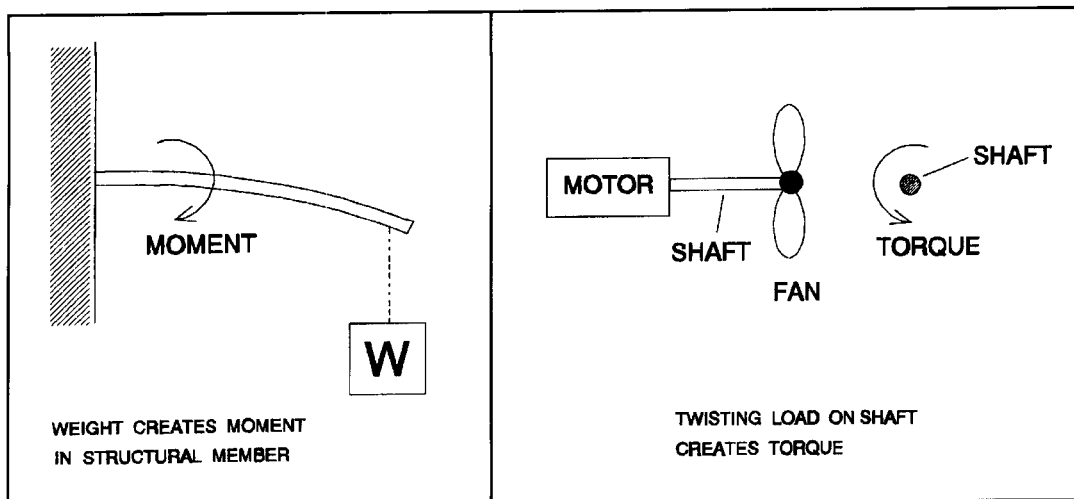
Figure 3. The difference between torque and moment.

mitted from other structural members, or combinations of these factors. Moments are transmitted down the length of a member, but they are not trying to rotate the member about its axis. Moments cause the bending of a member in the two planes perpendicular to its axis.[6] In amateur radio literature, torque and moment are often mistakenly interchanged, probably because they have the same units of measure. Also, incorrect units are commonly used. The proper units are "length-force." Examples of these are "inch-pounds" and "meter-kilograms." Although incorrect, it is quite common to see rotor torque listed as kg/m, which reads kilogram per meter. **Figure 3** further illustrates the difference between torque and moment.

Mast torque

Torque is produced when a force acts on a member that is attached to some form of shaft or axial member. It is produced by a force on a lever, or Yagi boom, as shown in **Figure 4**. For a Yagi, when wind forces due to antenna components are applied to the boom, a torque is produced that tries to turn the mast. Torque results in the mast because it is an axial member, while moments are produced that bend the boom. The only effect of interest for balancing the antenna is the net torque at the mast. For there to be zero net torque, the torque produced by antenna components to one side of the mast must be equal and opposite to that produced by the antenna components on the other side of the mast. In a perfectly balanced Yagi, the net torque would be zero, although there could be extremely high bending moments in the boom on each side of the mast.

In strict mathematical terms, torque is found by determining the vector cross product of the radius vector and the force vector that describes the magnitude and direction of the lever arm and the applied force. This is shown in **Equation 3**.[7]

$$\overline{T} = \overline{R} \ X \ \overline{F} \qquad (3)$$

$\overline{T}$ = torque vector

$\overline{R}$ = radius vector

$\overline{F}$ = force vector

Vector **Equation 3** can be simplified to scalar **Equation 4**.[8] In **Equation 4**, R and F are the magnitudes of the radius and force, while sin(z) is the sine of the angle, z, between the directions of the radius and force vectors. (The vector cross product can be looked at in two ways. It can be viewed as finding the magni-
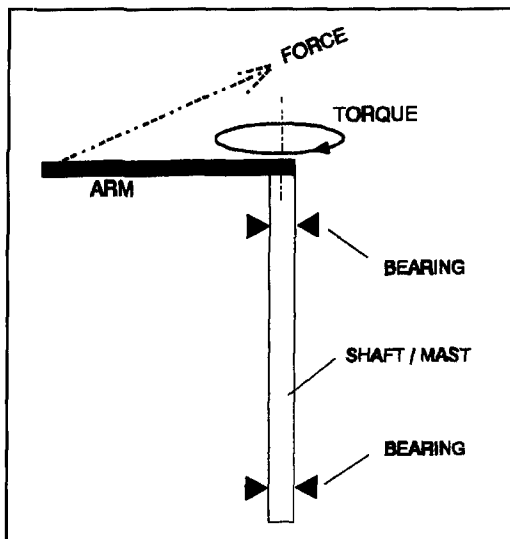
Figure 4. Force applied to a lever arm creates torque.

tude of R's component, which is perpendicular to the direction of F, and multiplying the component's and F's magnitudes while assigning the proper vector direction to T based on the "right hand rule." Or, it can be viewed as finding the magnitude of F's component that is perpendicular to the direction of R, and multiplying the component's and R's magnitudes while assigning the proper vector direction to T based on the "right hand rule." I will make use of the second view here; later in this article, I will use the first view. Mathematically, both views are identical.)

$$T = R \ F \ sin(z) \qquad (4)$$

T = torque
R = radius
F = force
z = angle between R and F

An example of **Equation 4** appears in **Figure 5**. In this figure, the radius is the distance from the mast to the point of application of force, F. **Equation 4** is easier to understand if you look at the term F sin(z). F sin(z) is the component of F perpendicular to the radius, or boom, as shown in **Figure 6**. **Equations 3** and **4** state that it is the component of F perpendicular to the boom that creates torque. In **Figure 6**, the other component of F is in line with the boom and creates no torque. This is easy to understand because the in-line force component results in tension and no twisting effect

To further understand the generation of torque, two special examples are shown in **Figures 7** and **8**. In **Figure 7**, the force is in line with the boom. As a result, no torque is created. In **Figure 8**, the force is perpendicular to the boom. In this case, the full magnitude of F creates torque.
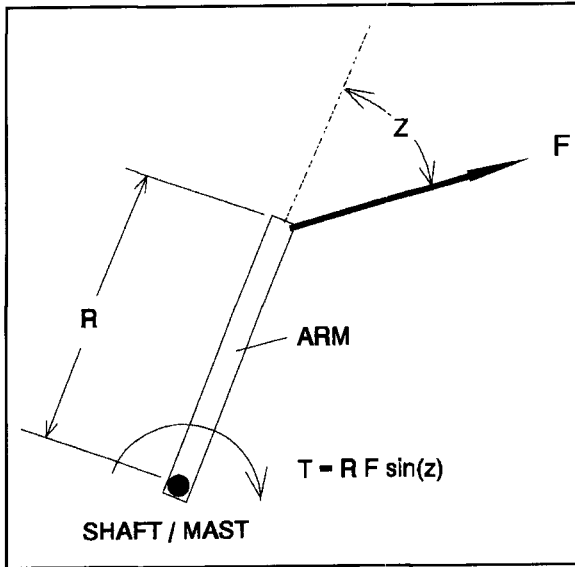
Figure 5. A pictorial representation of *Equation 4*.
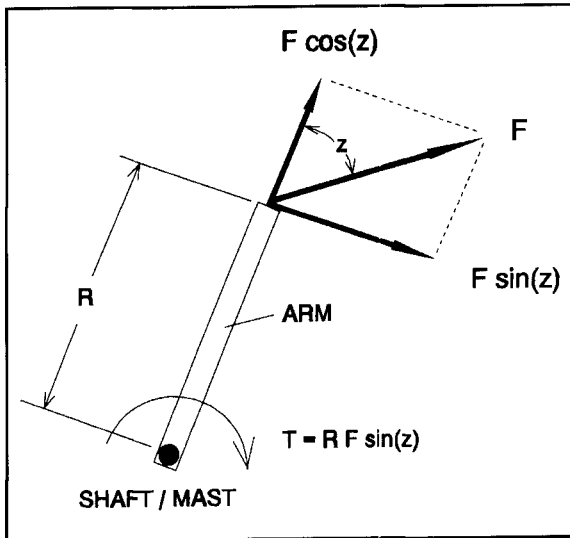
Figure 6. F sin(z) is the component of F perpendicular to the radius, or boom.

Rotating dipole and boom balancing

The simplest Yagi could be considered a rotating dipole. The act of balancing a simple dipole made of tubular members demonstrates many principles that will be used later for multi-element Yagis. In actuality, a Yagi boom and rotating dipole are identical in terms of wind balancing. For this case, assume the rotating dipole, or boom, is mounted on top of its mast and not to the side.

The wind force on a member is distributed and not applied at a single point. As a result, determination of the torque due to a distributed load requires the use of calculus to derive an equation in terms of wind loading and

Equation 4. For the simple case where the dipole (boom) half has the same diameter, the resulting torque is found using **Equation 5.** **Figure 9** shows a tube with a distributed wind loading and its equivalent obtained from **Equation 5.** In the equivalent, the force, F, found by **Equation 2,** is applied at the midpoint of the span, L/2, and is always perpendicular to the boom.

$$T = \frac{L\ F}{2} \qquad (5)$$

T = torque
L = length
F = net force on tube (lbs) (from **Equation 2**)

When the boom, or dipole, is made of tubes with different diameters, a generalized version of **Equation 5** can be derived. The generalized case is defined by **Equation 6** and an example of a three segment dipole-half is shown in **Figure 10.**

$$T = \frac{L1\ F1}{2} + \frac{(L2 + L1)\ F2}{2} + \frac{(L3 + L2)\ F3}{2} \qquad (6)$$

T = torque
F1 = force on segment 1
F2 = force on segment 2
F3 = force on segment 3
L1 = length, mast to segment end
L2 = length, mast to segment end
L3 = length, mast to segment end

Figure 11 shows a boom, or dipole, in a wind stream. The torque that results is the sum of the torque due to side 1 trying to turn the mast clockwise and the torque due to side 2 trying to turn the mast counterclockwise. Using the torque sign convention in **Figure 11,** the net torque produced by the boom is shown in **Equation 7.**

$$Tb = T1 - T2 \qquad (7)$$

Tb = net boom torque
T1 = torque due to side 1
T2 = torque due to side 2

For a rotating dipole, or boom, a torque balanced condition exists when T1 and T2 have the same magnitude, or the net torque is zero. Consider a single tube of total length L, attached to a mast as shown in **Figure 12.** At what point should the tube be attached to the mast to achieve a balanced condition, or have Tb = 0? For Tb = 0, T1 = T2. Forces Fa and Fb are obtained from **Equation 2,** and Ta and Tb are found using **Equation 5.**

$$Fa = C\ D\ La\ P\ \sin^2(ang)$$

$$Ta = \frac{C\ D\ La^2\ P\ sin^2(ang)}{2} \qquad \text{(clockwise)}$$

$$Fb = C\ D\ Lb\ P\ sin^2(ang)$$

$$Tb = \frac{C\ D\ Lb^2\ P\ sin^2(ang)}{2} \qquad \text{(counter-clockwise)}$$

For $Ta = Tb$,

$$\frac{C\ D\ La^2P\ sin^2(ang)}{2} = \frac{C\ D\ Lb^2P\ sin^2(ang)}{2}$$

after simplifying,

$$La = Lb$$

with $L = La + Lb$

$$La = Lb = \frac{L}{2} \qquad (8)$$

This result is not too surprising because it creates identical conditions on each side of the mast. For the case where there are three round segments on each side of the mast as shown in **Figure 13**, a balanced condition is described by **Equation 9**. (**Equation 9** can be used when there are three or fewer segments per side by setting the lengths and the diameters of nonexisting segments to zero.) Using **Equation 9**, an infinite number of combinations of diameters and lengths can be found that achieve balance. However, to have a balanced Yagi boom or rotating dipole, the easiest solution is an arrangement that is identical on both sides of the mast.

$$L1^2D1 + (L2+L1)(L2-L1)\ D2 + (L3+L2)$$
$$(L3-\ L2)\ D3 =(L6+L5)(L6-L5)\ D6 +$$
$$(L5+L4)(L5-L4)\ D5 + L4^2D4 \qquad (9)$$

The balancing method described here only applies to members that are either at rest or turning at a very slow rate—like that achieved with a typical antenna rotor. High rotational speeds alter the balance by creating different effective wind velocities for the part of the boom turning into the wind and the part going with the wind. With different velocities, the dynamic wind pressures for each side of the boom are not the same.

Torque due to element wind loads

Earlier in **Figure 2**, it was shown that the resulting force on a tube while in a wind stream is always perpendicular to the axis of the tube.[1]
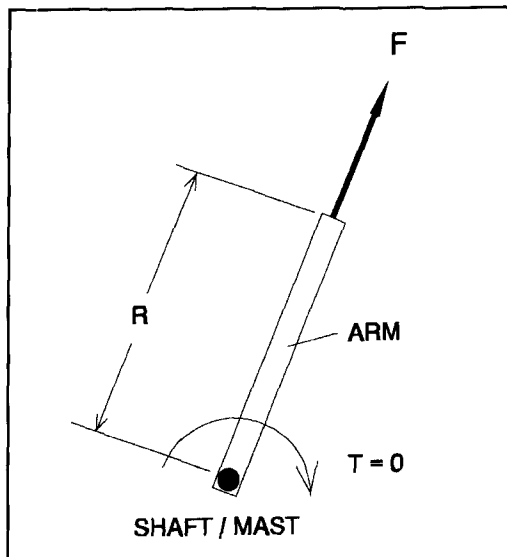
Figure 7. An example illustrating the generation of torque, in which the force is in-line with the boom. In this instance, no torque is created.
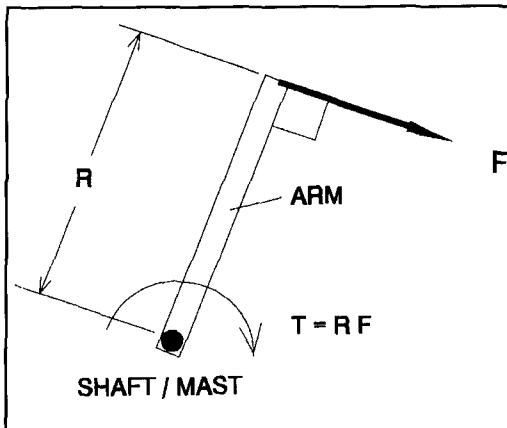
Figure 8. Another example illustrating the generation of torque. In this situation, the force is perpendicular to the boom, and the full magnitude of F creates torque.

When slightly drooped horizontal elements are mounted on a boom, the element forces are directly in line with the boom as shown in **Figure 14** for a single-element and **Figure 15** for a four-element Yagi. With element forces in line with the boom, there ar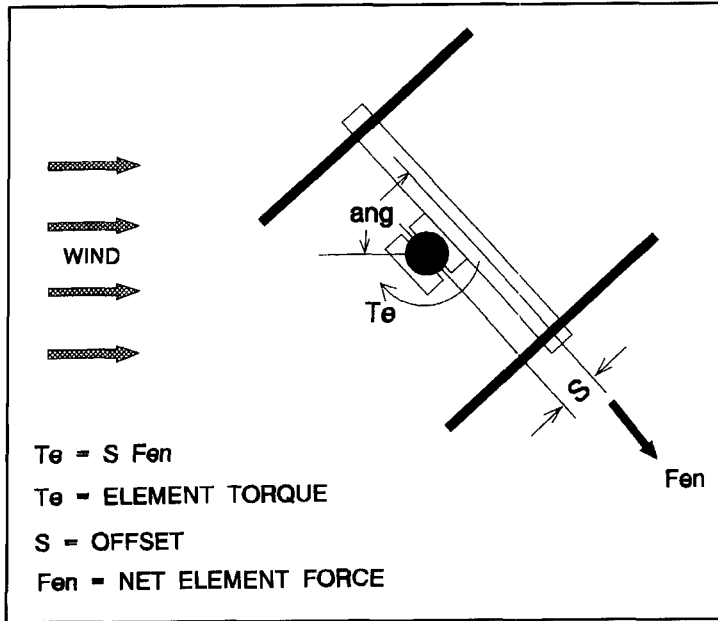e no large torques created by horizontally mounted elements. There is, however, a small torque due to element forces acting in conjunction with the boom-to-mast offset created by the boom-to-mast clamping assembly as shown in **Figure 16**. The resulting torque is created by the sum of the element forces down the boom, Fen, in combination with the boom-to-mast offset, S. **Equation 10** shows how to find Fen and **Equation 11** is used to find the torque, Te. When **Equations 10** and **11** are combined, a more useful form results as is shown in **Equation 12**.
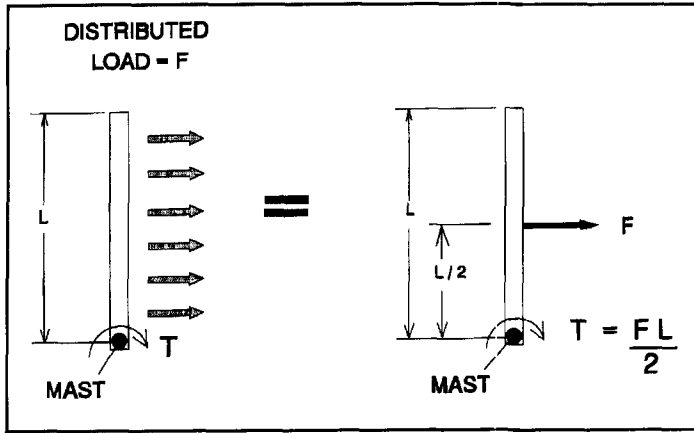
Figure 9. Equivalent loading: a tube with a distributed wind loading.

$$Fen = C\ Ae\ P\ cos^2(ang) \qquad (10)$$

Fen = element net force
Ae = net element projected area (sum of individual element projected areas)

The resulting torque is found using **Equation 12**, where S is the boom-to-mast offset.

$$Te = S\ Fen \qquad (11)$$

$$Te = C\ S\ Ae\ P\ cos^2(ang) \qquad (12)$$

Te = torque due to elements
S = offset

Equation 12 holds for elements with relatively small amounts of droop, like those found on Yagis built for 10 MHz and higher, and shortened elements used for 7 MHz. Full-size Yagi elements for 7 and 3.5 MHz have significant element droop, which results in a more complex loading situation. However, for the majority of Yagis used by amateurs, **Equation 12** is valid.

Offset boom torque

Earlier, I discussed balancing a boom when its center of rotation was directly in line with the axis of the mast. In this case, there was no offset, S, as shown in **Figure 16**. Because this mounting arrangement is not very common, boom torque must be addressed when there is an offset. **Figure 17** shows the boom force, Fb, on one side of the mast, at its point of effective application. In this figure, Rs is the radius vector used with Fb to find the torque for this instance. Using the vector cross product equation for torque as defined in **Equation 3** for this case, results in **Equation 13**. **Equation 14** is produced by changing **Equation 13** to its scalar form, as was done to obtain **Equation 4**.

$$\overline{Tbe} = \overline{Rs}\ X\ \overline{Fb} \qquad (13)$$

$\overline{Tbe}$ = boom torque vector
$\overline{Rs}$ = radius vector to Fb
$\overline{Fb}$ = boom net force vector

$$Tbe = Rs\ Fb\ sin(z) \qquad (14)$$

Tbe = boom torque due to offset
R = radius
Fb = boom net force
z = angle between Rs and Fb

From **Equation 14** the terms "Rs sin(z)" can be replaced by Re, which is the component of Rs perpendicular to the direction of Fb. This
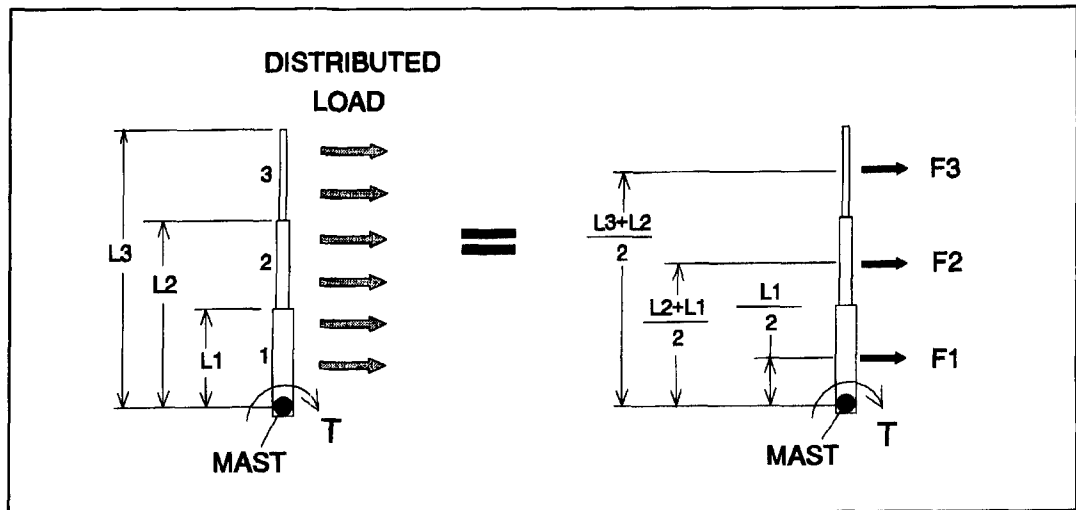
Figure 10. Equivalent loading: a three segment half-dipole.
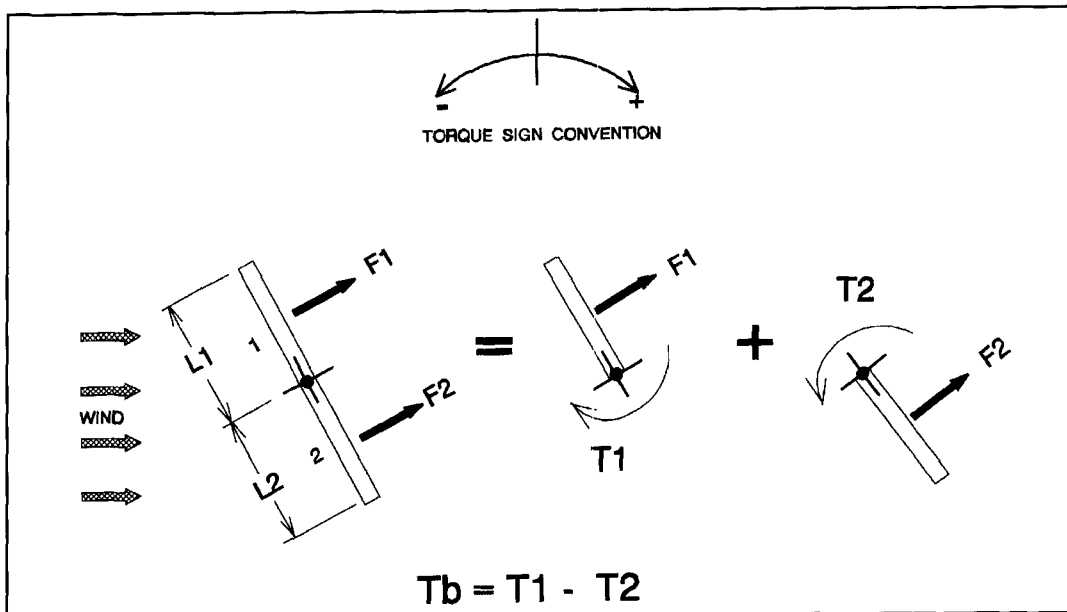
Figure 11. A boom, or dipole, in a wind stream.

results in a very convenient relationship, because Re has the length L/2. Or,

$$Tbe = Re \; Fb$$

where

$$Re = Rs \; \sin(z)$$

Since:

$$Re = \frac{L}{2} \qquad (15)$$

then

$$Tbe = \frac{L \; Fb}{2} \qquad (16)$$

Note that **Equation 16** is the same as **Equation 5**, which was found for the case where there was no boom offset. This means the offset has no effect on torque due to the boom. As a result, balanced conditions defined by **Equation 9** hold for booms with offsets. This greatly simplifies matters. To balance a boom, satisfy **Equation 9**. Or, for a more practical approach, have the boom construction to the left of the mast be exactly the same as that to the right of the mast.

Imbalance example

To begin, let's take a look at a large Yagi and observe the effects of the boom and the elements on its balance. **Figure 18** illustrates the approximate physical description of the well-known 20-meter Yagi—the KLM 20M-6 "Big Sticker." One thing to note is that the boom is

52 inches longer on one side of the mast. Because of this, you can expect a significant imbalance due to the boom.

Because there are two main contributors to a Yagi's torque imbalance, it is important to consider each contributor and the combined net effect. By looking at each individual contributor, it may be possible to make a design change to minimize its effect. First, find the torque due to the boom using **Equations 2, 5,** and **7** for sides 1 and 2.

$$F1 = C \; A1 \; P \; \sin^2(ang) = C \; D \; L1 \; P \; \sin^2(ang) \qquad (17)$$

$$F2 = C \; A2 \; P \; \sin^2(ang) = C \; D \; L2 \; P \; \sin^2(ang) \qquad (18)$$

F1 = side 1 boom force
F2 = side 2 boom force
C = 1.2 (drag coefficient for round boom)
D = boom diameter
P = dynamic wind pressure
L1 = side 1 boom length
L2 = side 2 boom length

Then, use F1 and F2 to find the torques:

$$T1 = \frac{L1 \; F1}{2} = \frac{1.2 \; D \; L1^2 P \; \sin^2(ang)}{2} \qquad (19)$$

$$T2 = \frac{L2 \; F2}{2} = \frac{1.2 \; D \; L2^2 P \; \sin^2(ang)}{2} \qquad (20)$$

T1 = side 1 torque

T2 = side 2 torque
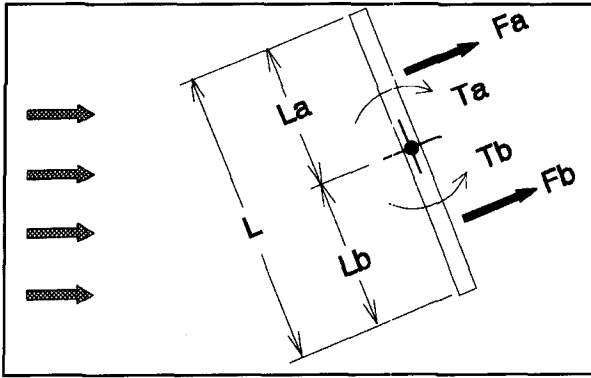
Figure 12. A dipole/boom balancing example.

Using the geometry from **Figure 18**, the parameters in **Equations 22, 23,** and **24** can be set as shown below. To get some idea of the magnitude of these effects, set the wind speed at 50 mph, which has a dynamic wind pressure of 6.4 psf. I calculated values of Tb and Te using **Equations 21** and **22** and plotted them in **Graph 1**.

$$P = 6.4 \text{ psf}$$
$$D = 3"$$
$$L1 = 320"$$
$$L2 = 372"$$
$$S = 4"$$
$$Ae = 12.6 \text{ sqft}$$

$$Tb = T1 - T2 = \frac{1.2 \ D \ P \ (L1^2 - L2^2) \sin^2(\text{ang})}{2}$$

$$(21)$$

Tb = resulting torque due to boom

To find the torque due to the elements, use **Equation 12**. With C = 1.2 for round elements you will find:

$$Te = 1.2 \ S \ Ae \ P \cos^2(\text{ang}) \qquad (22)$$

Te = torque due to elements
S = offset
Ae = net element projected area (sum of individual element projected areas)

With both Tb (boom torque) and Te (element torque) defined, the resulting net torque can be found by summing:

$$Tn = Tb + Te \qquad (23)$$

After substituting and combining, you get **Equation 24**, which describes the net torque due to the boom and elements as a function of "ang" as shown in **Figure 19**.

$$Tn = 1.2 \ P \ [\ 0.5 \ D \ (L1^2 - L2^2) \sin^2(\text{ang}) +$$

$$S \ Ae \cos^2(\text{ang}) \ ] \qquad (24)$$

Tn = net torque

Graph 1 shows that the mast torque due to the boom is much larger than the torque due to the elements. As can be seen in **Graph 2**, which has Tn plotted along with Tb and Te, the element torque is a minor contributor to the net torque for this Yagi configuration. One thing interesting to note from **Equation 12** for Te is that the position of the elements on the boom does not matter. It is the offset, S, and the net element projected area that affects Te.

The example just shown is a Yagi that can easily have its imbalance reduced. One need only make the length of the boom to the left of the mast equal to the length of the boom to the right of the mast. As a first step, suppose the boom-to-mast mounting plate assembly is moved so the boom is only 12 inches longer on one side rather than the original 52 inches. Using **Equation 24** with L1 = 340" and L2 = 352", new values of Tn are found and plotted in **Graph 3** along with Tn before boom modification. The improvement is quite dramatic for doing nothing more than moving the boom-to-mast plate. Looking at **Graph 4**, a plot of Te and Tb for the example with the altered boom-to-mast mounting plate, you will note that, even with a small difference between L1 and L2, the torque due to the boom is still higher than that due to the elements.

If you want a simple way to bring a Yagi into better balance, the first thing you should do is put the boom-to-mast mounting in the right
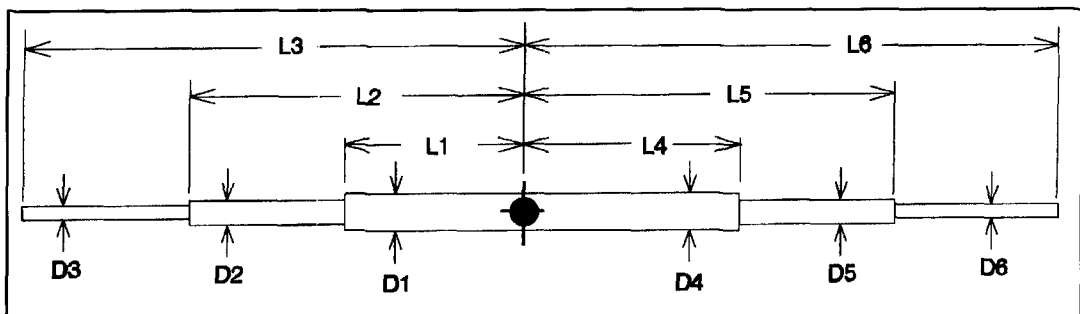
Figure 13. In this situation, where there are three round segments on each side of the mast, a balanced condition is described by *Equation 9*.

location on the boom. For a boom that has the same diameter over its entire length, the boom-to-mast mounting plate should be attached at the midpoint of its length. For a boom with different diameters, **Equation 9** should be satisfied—or, the boom should be the same on both sides of the mast.

When most Yagis are assembled, the boom-to-mast mounting plate is usually mounted at the weight balance point. However, this point is not normally coincident with the aerodynamic boom balance point. When the boom-to-mast mounting plate is at the best location for aerodynamic balance, a weight balance can be achieved by mounting a weight inside the light end of the boom. The addition of a weight will increase the Yagi's overall heaviness, but the achievement of a better aerodynamic balance will be well worth the effort. To obtain the best boom balance for a 125-pound KLM 20M-6, the boom-to-mast mounting plate should be moved 26 inches toward the last two directors. To bring the antenna back into weight balance, a weight of approximately 9.4 pounds should be mounted inside the director end of the boom. In this instance, 9.4 pounds is a small price to pay to reduce the imbalance.

Alternative boom balancing methods

There is another way to balance a boom. This method requires that a flat sheet be mounted to the boom. It can be made from a thick piece of aluminum sheet metal U-bolted to the "short" side of the boom, as shown in **Figure 20**. The goal is to bring the boom into balance by creating more torque on the side of the boom that produces the lesser torque before modification. The size and location of the plate are determined with an eye to equalizing the torques produced by both sides of the boom.

Use **Equation 21** to find the boom torque imbalance before modification. Once the initial imbalance is known, an equation can be written to determine the size of a torque balance plate as a function of its location on the boom. To begin, you first need an equation for the force due to the plate. Since the "cross flow principle" applies to flat surfaces as well, you can write a force equation similar to **Equation 2** for the plate. **Equation 25** results. With the plate located at a distance of Rpb from the mast, a torque results that can be determined using **Equation 26**. In actuality, the plate blocks part of the boom from the wind. Blockage is an effect which will be accounted for in a later step.

$$Fp = Cp \, Ape \, P \, sin^2(ang) \qquad (25)$$
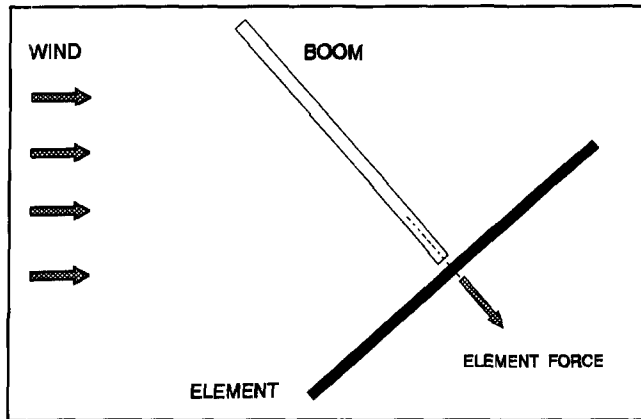
Figure 14. When slightly drooped horizontal elements are mounted to a boom, the element forces are directly in line with the boom, as shown here for a single-element Yagi.
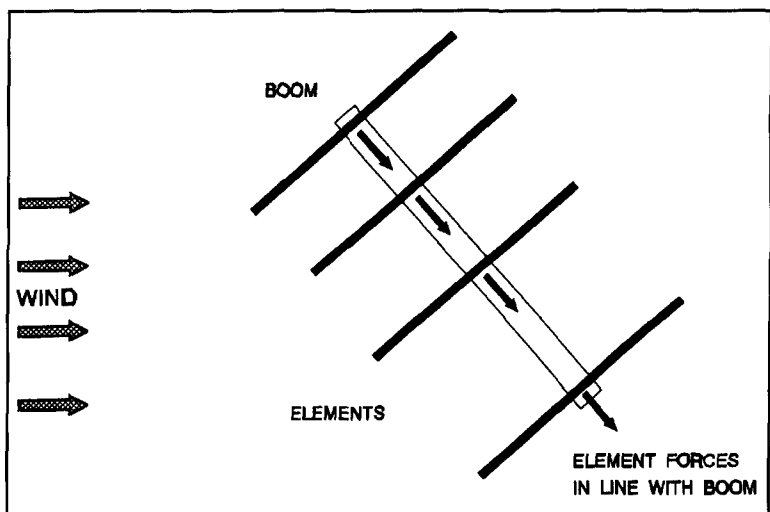
Figure 15. The same situation described in *Figure 14*, for a four-element Yagi.

Fp = plate wind force
Cp = plate drag coefficient
Ape = effective plate projected area

$$Tp = Rpb \, Fp \qquad (26)$$

Tp = torque due to plate
Rpb = distance from mast axis to centerline of plate

Substituting Fp from **Equation 25** into **Equation 26** produces **Equation 27**, which now includes Ape and Rpb.

$$Tp = Rpb \, Cp \, Ape \, P \, sin^2(ang) \qquad (27)$$

The next step is to set the torque due to the boom balance plate (Tp), plus that due to the boom on side 2 (T2), equal to the torque due to the boom on side 1 (T1). The imbalance between sides 1 and 2 will be made up by the

Figure 16. There is a small torque due to element forces acting in conjunction with the boom-to-mast offset created by the boom-to-mast clamping assembly shown here.

torque created by the boom balance plate. Torque balance is achieved when **Equation 28** is satisfied.

$$T1 = Tp + T2 \qquad (28)$$

Using **Equation 19** for T1, **Equation 27** for Tp, and **Equation 20** for T2 results in **Equation 29**. This can be simplified and solved for Ape to provide **Equation 30**.
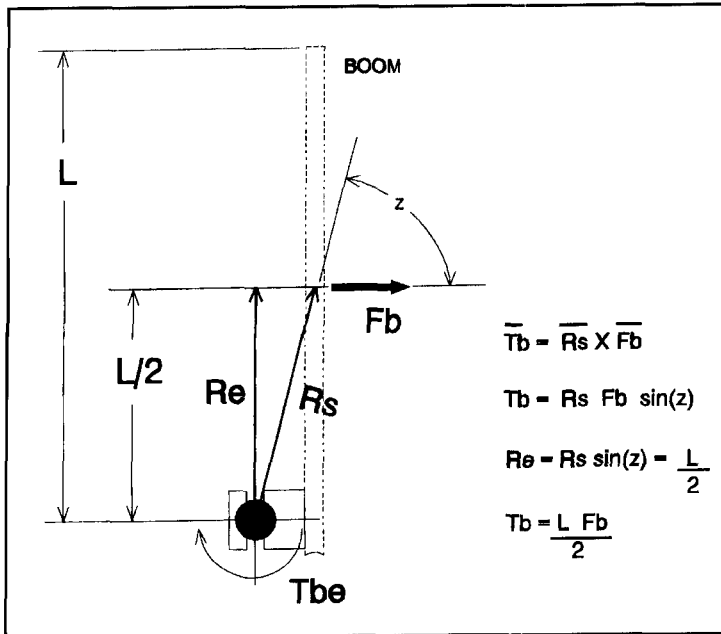
Figure 17. The boom force, Fb, on one side of the mast, at its point of effective application.

$$\frac{1.2 \, D \, L1^2 \, P \, \sin^2(ang)}{2} =$$

$$Rpb \, Cp \, Ape \, P \, \sin^2(ang) +$$

$$\frac{1.2 \, D \, L2^2 \, P \, \sin^2(ang)}{2} \qquad (29)$$

$$Ape = \frac{1.2 \, D \, (L1^2 - L2^2)}{2 \, Rpb \, Cp} \qquad (30)$$

Also, since the balance plate you are using is square, the plate drag coefficient, Cp, can be set equal to 1.4.[9]

$$Ape = \frac{0.43 \, D \, (L1^2 - L2^2)}{Rpb} \qquad (31)$$

When **Equation 30** or **31** is solved for Ape, a negative value may result. As **Equation 28** is written with the plate on side 2, a negative value of Ape indicates the plate should be mounted on side 1 at a projected distance Rpb from the mast. Once you have Ape from **Equation 31**, you can find the true dimensions of a square torque balancing plate of size W x W that accounts for blockage of the boom. Dimension W can be found using **Equation 31**, where the absolute value of Ape is used.

$$W = 0.43 \, D + [\, 0.184 \, D^2 + ABS(Ape) \,]^{0.5} \qquad (32)$$

Using the KLM 20M-6 again as an example, with the torque balancing plate at Rpb = 23' 7" as shown in **Figure 20**, you can calculate Ape from **Equation 31**. With L1 = 320, L2 = 372, and D = 3", Ape is found to be -1.134 sqft. Because Ape has a negative value, the plate should be mounted on side 1 as shown in **Figure 20**. Knowing Ape, **Equation 32** can be used to adjust for the boom blockage. The resulting value for W is 1.18 ft for an actual plate area of 1.39 sqft.

Let's look at a third way to balance a boom. In the case of a boom that has the same diameter along its length, you are attempting to counteract the difference in boom lengths, side 1 as compared to side 2. A simple way to balance the boom is to add a boom extension to the shorter side as shown in **Figure 21** to make it the same length as side 2. For stepped booms as shown in **Figure 13**, any addition to the boom has to meet the requirements of **Equation 9**.

Boom balancing summary

The examples showed that a small amount of boom asymmetry can result in large mast torques—even larger than the contribution due
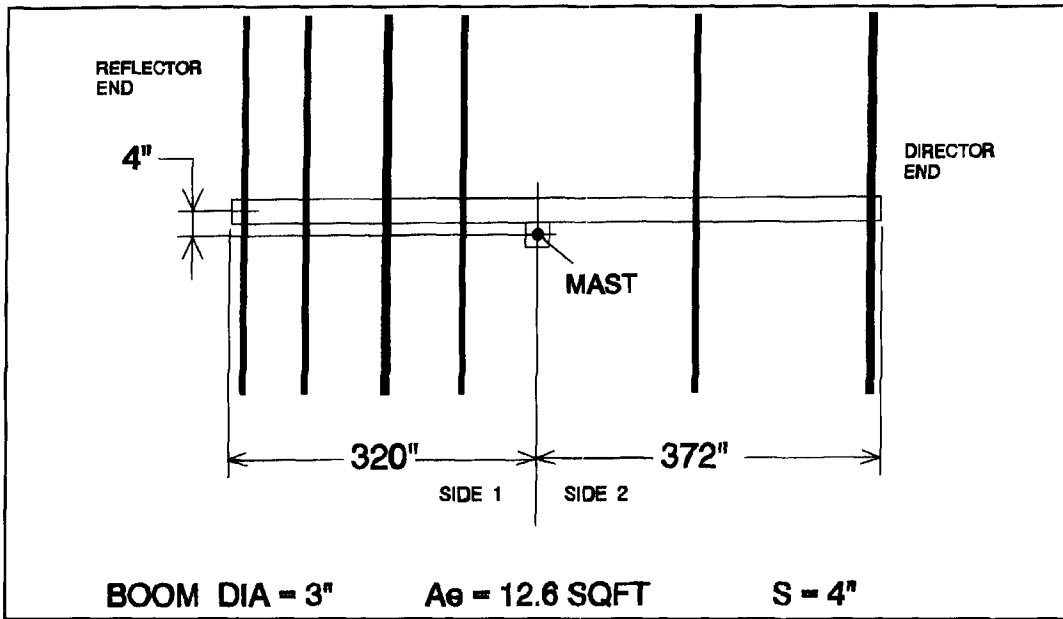
Figure 18. Approximate physical description of the KLM 20M-6 "Big Sticker."

to the elements. Consequently, efforts to bring a Yagi into better balance should first be focused on the boom. For a constant diameter boom, the boom-to-mast mounting plate should be located at the boom midpoint. For a stepped boom, it is easiest to have side 1 and side 2 identical. If not, **Equation 9** should be satisfied. In each of these cases it is most probable

that a weight will need to be mounted inside one end of the boom to bring the assembled Yagi into weight balance.

Alternative approaches to balancing a boom include adding a torque balance plate or a boom extension. For a Yagi already installed on a tower, the best approach may be to add a torque balance plate. The additional weight will
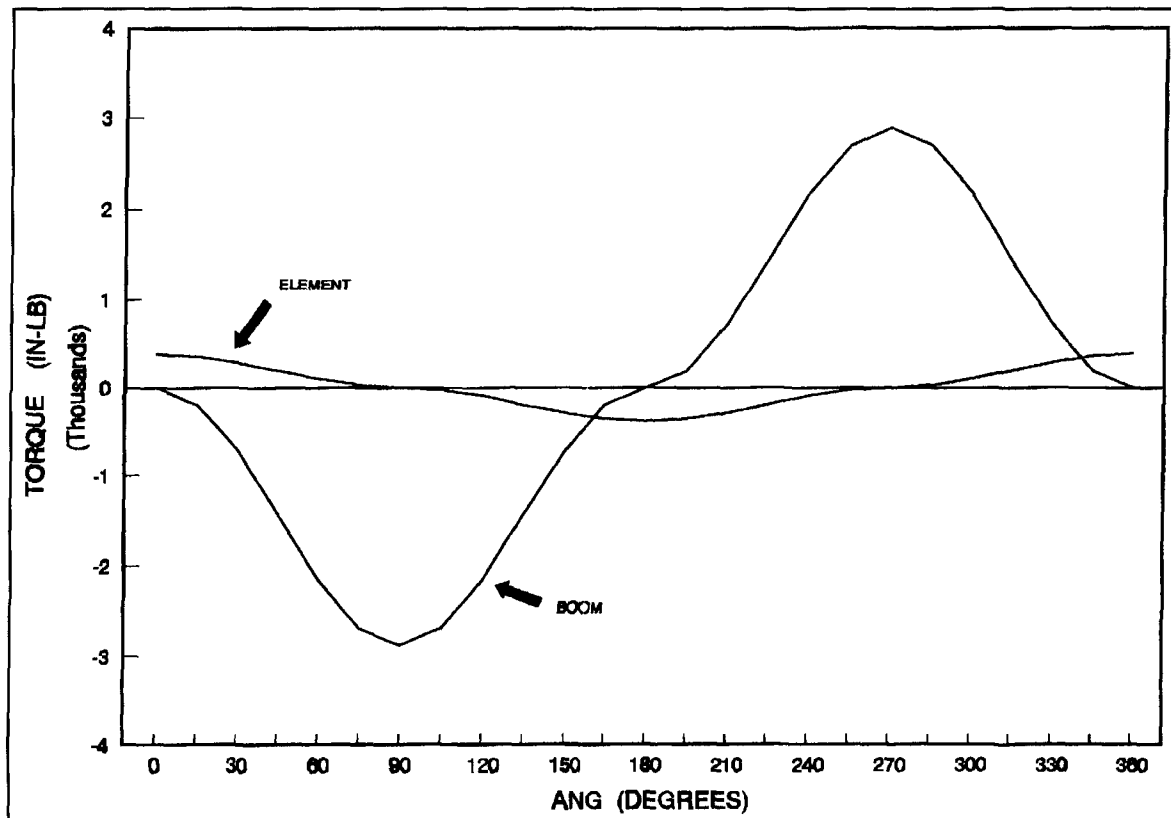
Graph 1. Element and boom torque.
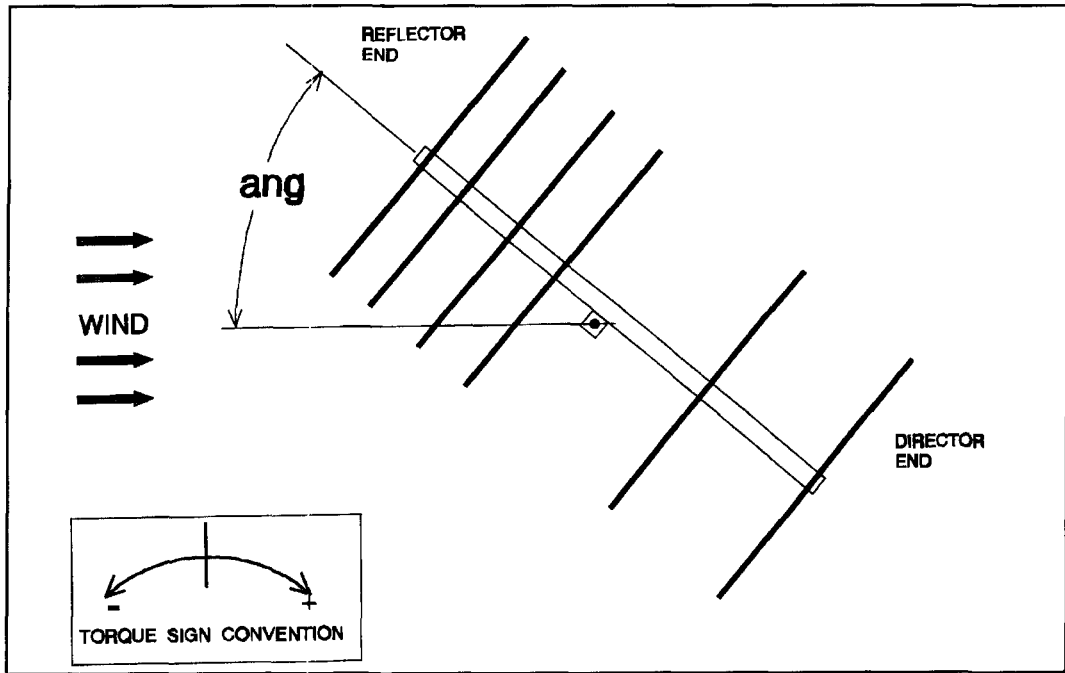
Figure 19. The net torque due to the boom and elements as a function of "ang".

be negligible and, most likely, not appreciably alter the existing weight balance.

In part 2

In part 2 of this article, I will continue my discussion of aerodynamic balancing by addressing imbalance due to elements, element torque balancing, the design and use of an off-set mast, and side arm mounted Yagis. I will also summarize information discussed in both parts 1 and 2.
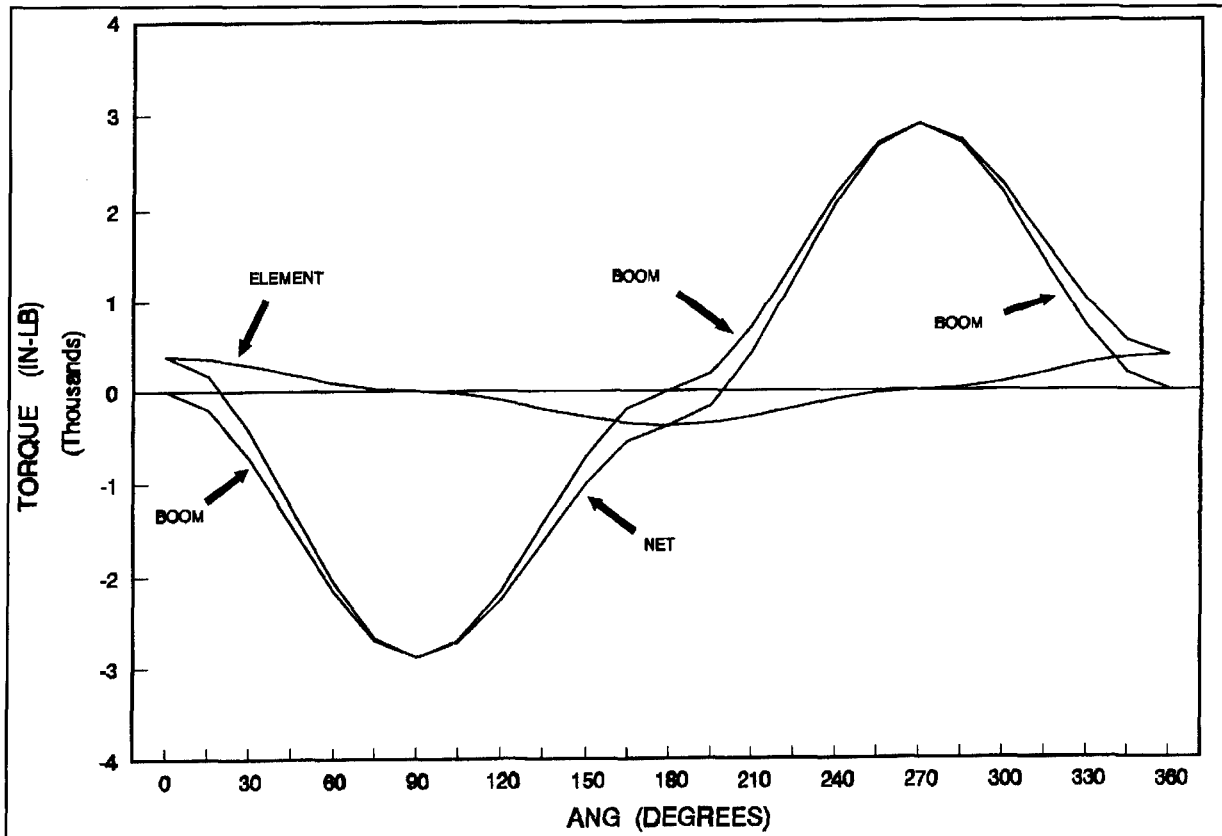
Graph 2. Element, boom, and net torque.
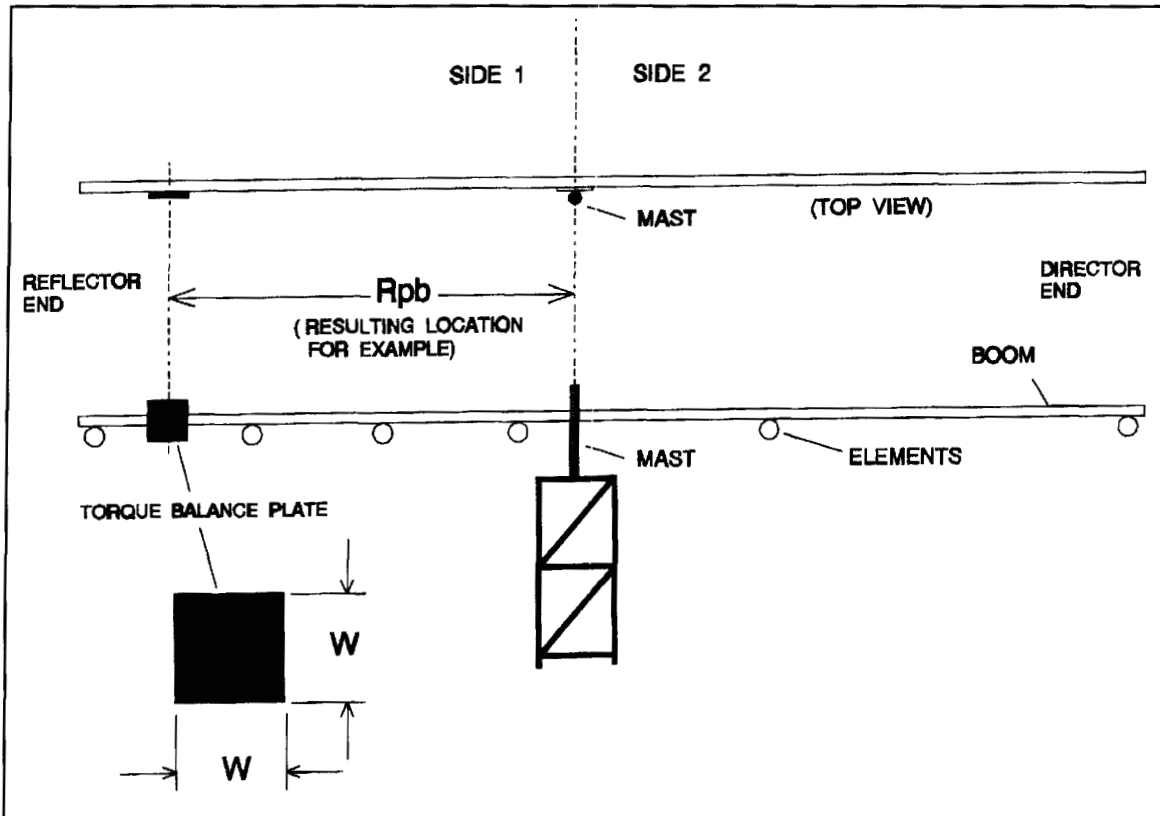
Figure 20. To balance a boom, you can U-bolt a thick piece of aluminum sheet metal to the "short" side of the boom..
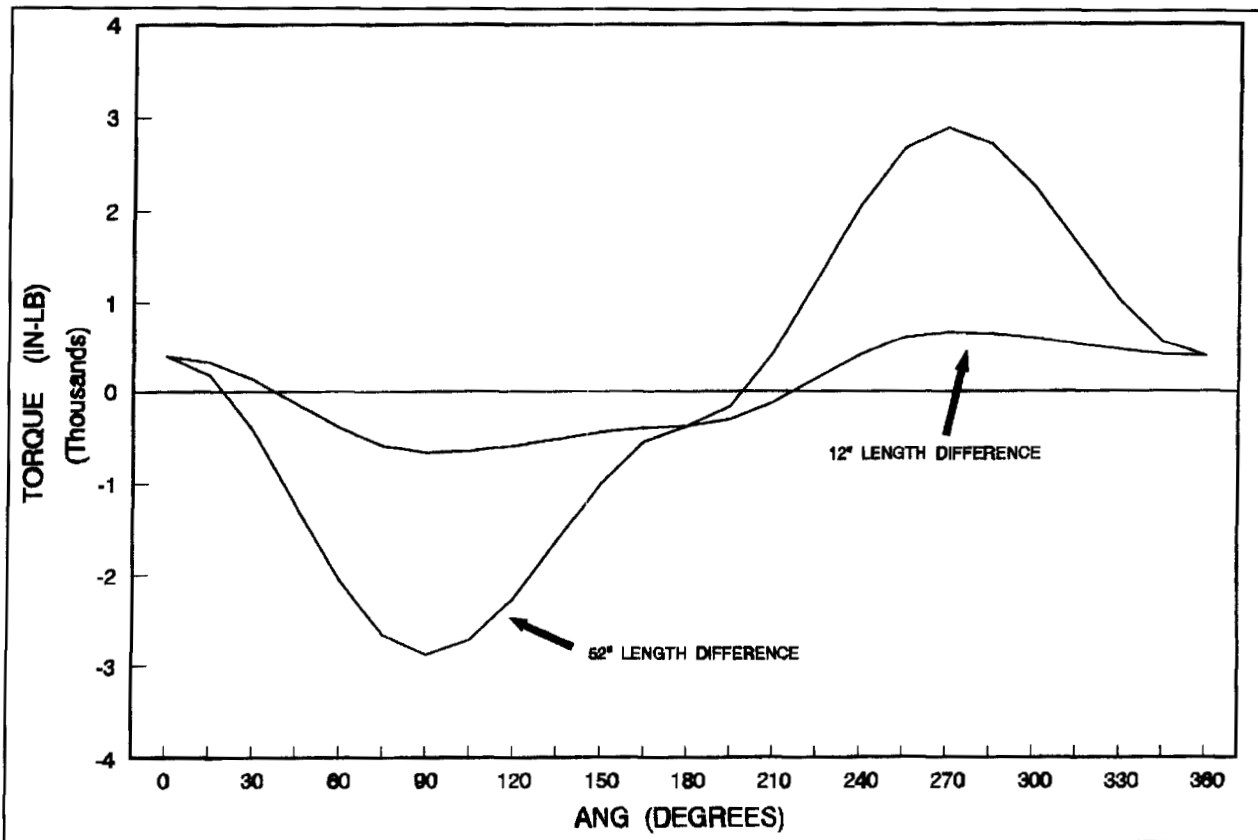
Graph 3. Net torque.
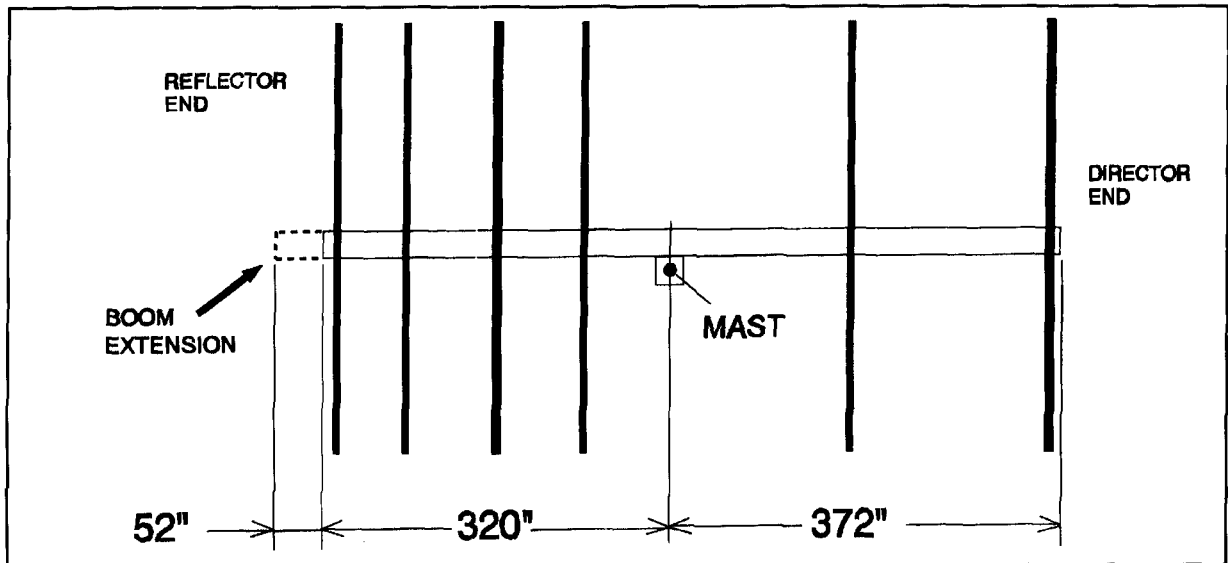
Figure 21. A simple way to balance the boom is to add a boom extension to the shorter side to make it the same length as side 2.

REFERENCES
1. Dick Weber, K5IU, "Determination of Yagi Wind Loads Using the Cross Flow Principle," Communications Quarterly, Hicksville, New York, Spring 1993, pages 13-30.
2. Roger Cox, WB0DGF, "Determining Antenna Surface Area," Telex-Hygain, Minneapolis, Minnesota, 1978, page 4.
3. David B. Leeson, W6QHS, Physical Design of Yagi Antennas, American Radio Relay League, Newington, Connecticut, 1992, page 6-1.
4. S.F. Hoerner, "Fluid-Dynamic Drag," published by author, Bricktown, New Jersey, 1965, page 3-11.
5. Dick Weber, K5IU, "Determination of Yagi Wind Loads Using the Cross Flow Principle," Communications Quarterly, Hicksville, New York, Spring 1993, page 15.
6. J.L. Meriam, Engineering Mechanics, Statics and Dynamics, John Wiley & Sons, New York, New York, 1980, page 217.
7. George W. Housner and Donald E. Hudson, Applied Mechanics, Statics, D. Van Nostrand Company, Princeton, New Jersey, 1964, page 23.
8. Wilfred Kaplan, Advanced Calculus, Addison-Wesley Publishing Company, Reading, Massachusetts, 1959, page 47.
9. EIA/TIA-222-E, "Structural Standards for Steel Antenna Towers and Antenna Supporting Structures," Electronics Industries Association, Washington, D.C., 1991, page 8.
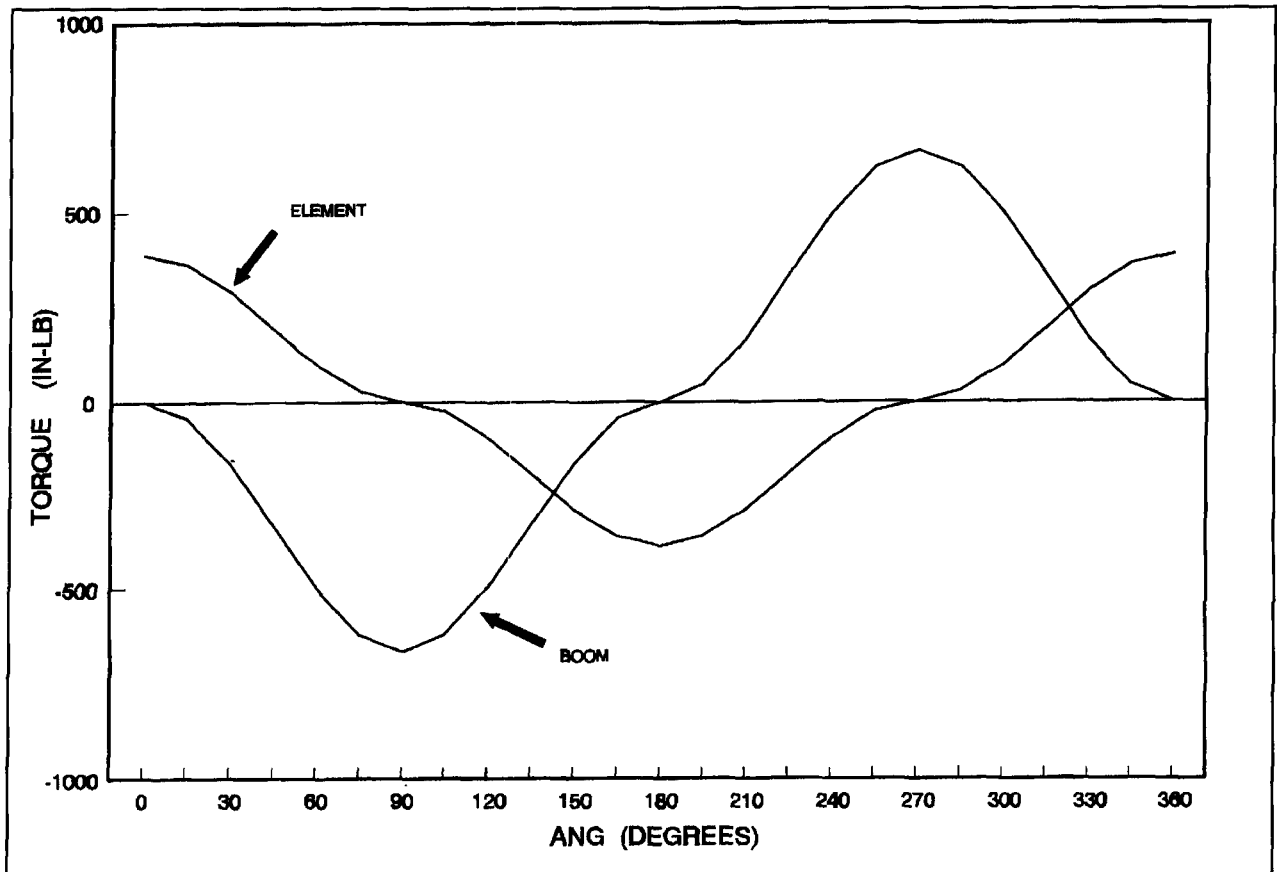
Graph 4. Plot of Te and Tb for example with altered boom-to-mast mounting plate

William E. Sabin, *WØIYH*
1400 Harold Drive
Cedar Rapids, Iowa 52403

DESIGNING THE LONG-WIRE ANTENNA SYSTEM

Want to design a simple antenna system? This article takes you through the steps.

A s part of my home laboratory and ham station, for wideband (10 kHz to 30 MHz) receiving and also for occasional operating on all the HF ham bands, I decided to use an end fed wire antenna about 140 feet (total) long but only about 25 feet above ground. I became interested in studying the electrical properties and system design of this antenna and, despite its very modest claims to elegance, in understanding and optimizing whatever capabilities it might have. I wasn't especially concerned with its directional properties, but the results of my computer simulations are described briefly in **Appendix A**. The procedures I used to design and analyze this system can be applied to many other specific
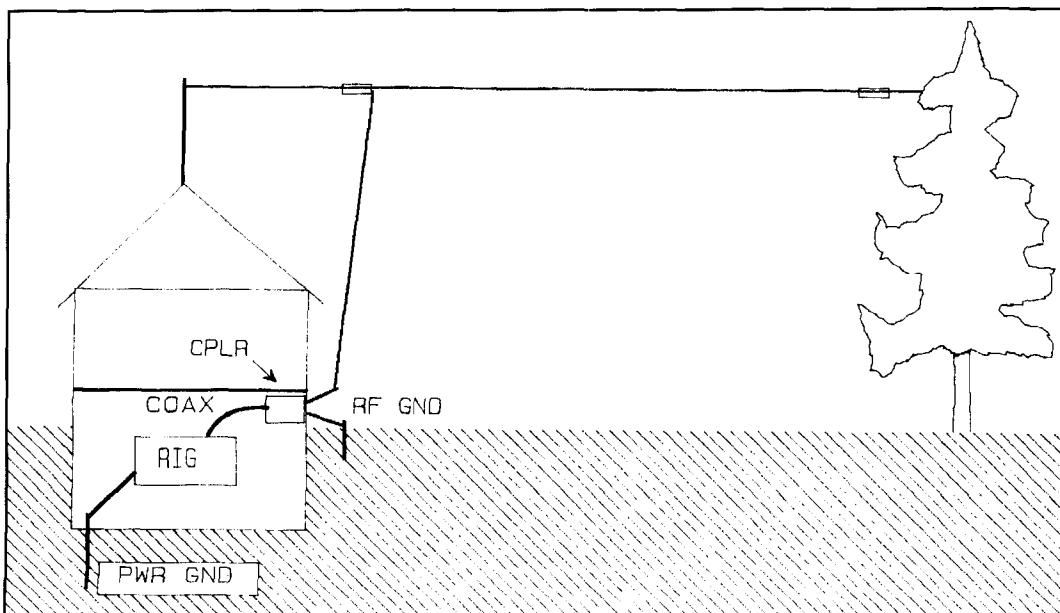
Figure 1. Arrangement of house, basement, antenna, RF ground, and power ground.
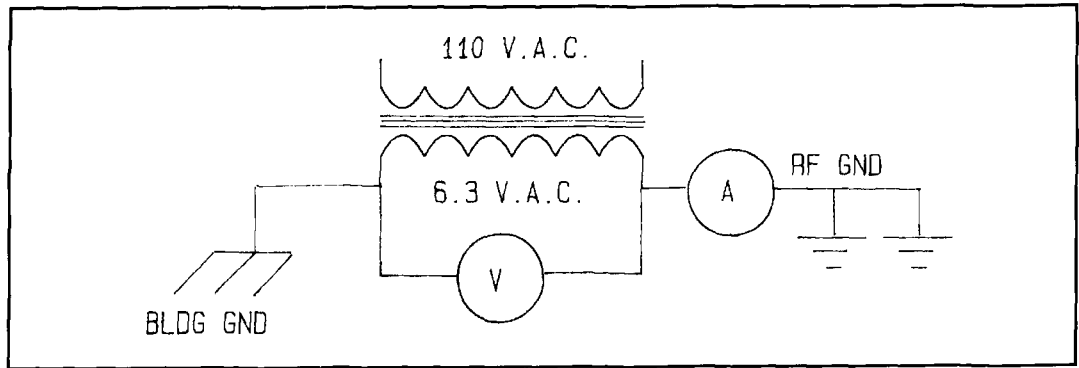
Figure 2. Measurement method for ground resistance.

situations, and may be of particular interest to those who need a simple and unobtrusive multi-band antenna or perhaps a temporary field day setup. My major goals were to obtain a 50-ohm interface with my equipment, reduce harmonic radiation, maximize radiation efficiency, and reduce transmitter RF inside the house. The absence of a heavy transmission line that must be supported by the wire and its end fastenings enhances the simplicity.

The end-fed antenna

Figure 1 shows the general layout and is pretty much self-explanatory. There are no towers or

supports except a 10-foot mast on my one-story house and a tree at the far end of my back yard. The end of the wire connects directly to the coupler (transmatch) network located right at the basement window, and the quality of the return path to ground is an important system consideration. A simple roped-off boundary outside the basement window keeps people and pets a safe distance from the wire. A DC path from antenna wire to RF and building grounds is always in place. Power output level is limited to 100 watts to reduce any possible physiological hazards from closely adjacent RF fields.

The horizontal end fed, or Hertz, antenna requires either a high quality ground connec-
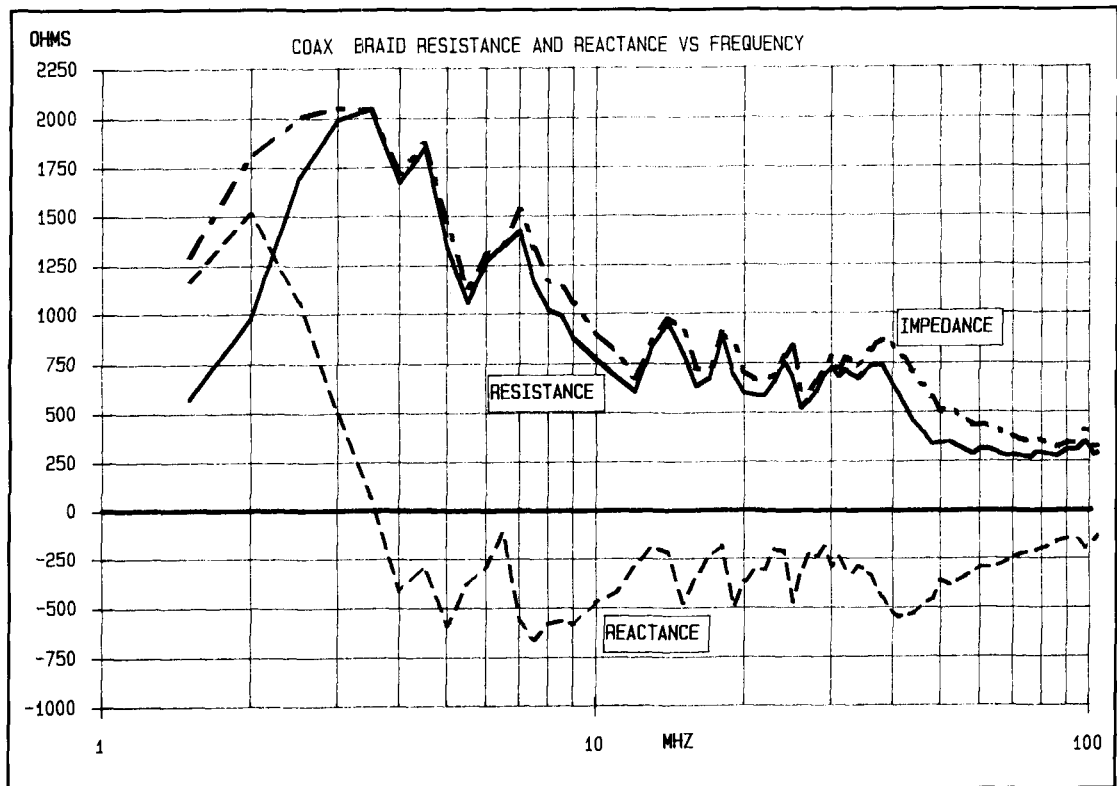
Figure 3. Series resistance, reactance, and impedance of a 15-foot ferrite-loaded RG-58 coax braid.
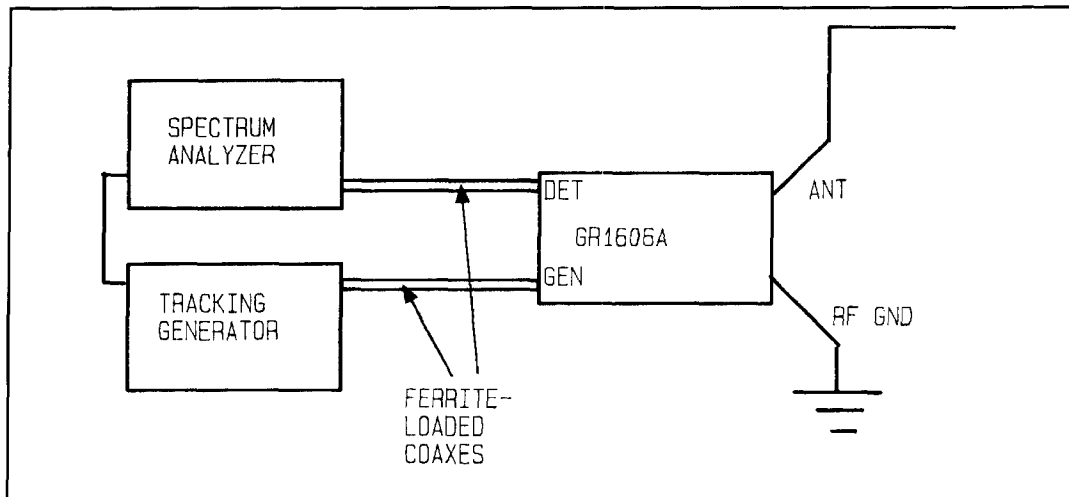
Figure 4. Test setup to measure antenna impedance.

tion or some other reference plane to "launch" a wave efficiently along its length, as compared to a balanced antenna in which each side can be thought of as a counterpoise for the other side. The true Zepp antenna, end fed, is a possible exception if the two-wire feedline is long enough; however, as the feedline length approaches zero wavelength, it too becomes less self-supporting and requires a local reference plane (for example, a Zeppelin). The reference plane for my antenna is, or should be, my back yard. In other words, any linear antenna is essentially a two-terminal device.

The balanced antenna is often preferable in terms of overall radiation efficiency, but in situations such as mine other electrical and even "backyard environmental" factors are involved, as mentioned at the beginning of this article, which require certain compromises. For me, the exercise was to obtain the best results within that framework.

Designing the ground system

In the ideal situation, the transmatch chassis is at true ground potential—even at the highest frequency. The antenna circuit includes the transmatch, the wire, the ground plane, and the open space between wire and ground. No RF currents flow from the transmatch into the equipment or other parts of the home environment. The induction field of the antenna is far removed from the "radio shack," and the ground plane is large and of zero impedance—therefore it doesn't absorb any radiation.

In **Figure 1**, the ground is less than ideal, the RF ground connection has significant wavelength and loss resistance, and is part of the antenna radiation system with three undesirable results. These are:

A) Some very inefficient radiation occurs in or near the basement, not out in the open.

B) Power is wasted in the ground resistance.

C) An RF potential difference exists between the coupler chassis and the radio equipment. Therefore the outside of the coax, the equipment, and the house wiring also become part of the antenna system.

At the higher frequencies, the increasing impedance of the preferred RF ground path aggravates these problems.

I used two methods to improve the effects of imperfect ground.

A) Improve the ground connection. I installed two 8-foot ground rods three feet apart, and connected each to the transmatch ground connection with 1 inch by 3 foot braided straps to minimize lead inductance. Using the method shown in **Figure 2**, I measured a total loop resistance of 38 ohms at 60 Hz using one ground rod and 22 ohms using two rods. From this data, I estimated the power ground at 6 ohms and the RF ground at 16 ohms. More ground rods would be marginally helpful. I didn't want to construct elaborate ground screen or counterpoise installations; but a number of ground wires, placed several inches below the level of the lawnmower blade and directly underneath the antenna wire, would be worthwhile where feasible. Still, the ground rods by themselves were a satisfactory compromise for me, and the overall design is such that the power lost is minimized. They also provide some lightning protection.

B) To isolate the antenna coupler from the equipment and to encourage the preferred RF ground path, I loaded the RG-58 antenna coax with Amidon FB-XX-6301 ferrite beads (remove the coax outer insulation). These either choke off or absorb RF that appears on the outside of the coax. The coax is approximately 15 feet long and the beads are distributed about 3 inches apart (this gave the best measured results). I alternated two kinds of beads, XX =

RESISTANCE VS FREQUENCY

REACTANCE VS FREQUENCY
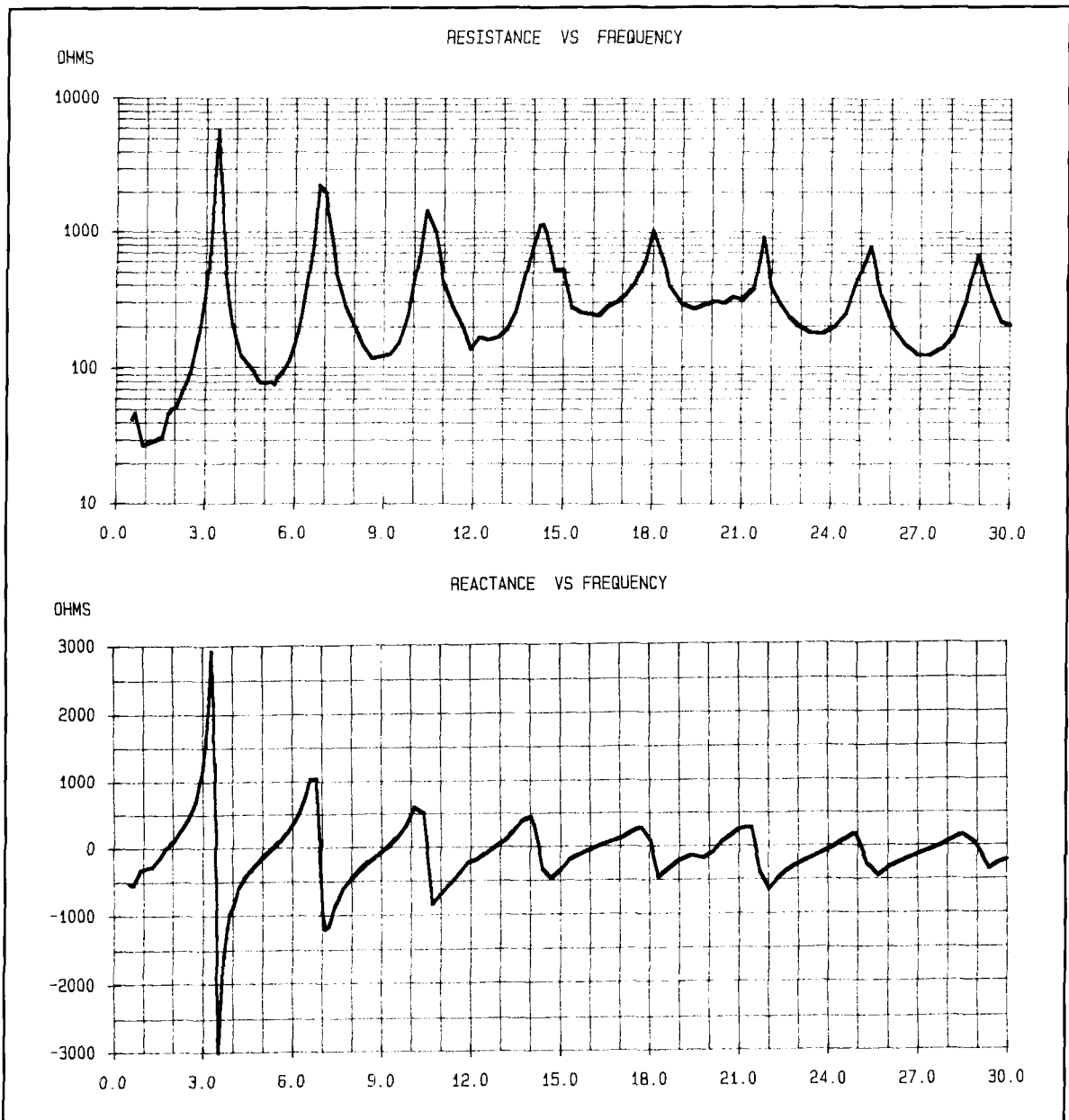
Figure 5. Resistance and reactance of my 140-foot long wire antenna, including the ground path.

43, μ of 850, and XX = 77, μ of 2500. Heavy vinyl tubing covers the finished cable. I used this approach to obtain isolation from 1.8 MHz to 30 MHz and the harmonic frequencies at VHF and UHF. It also breaks up a possible RF ground loop between building ground and antenna ground. **Figure 3** plots the linear resistance and reactance of the coax braid, as measured on my HP Vector Impedance Meter, up to its limit of 110 MHz. I like this distributed approach because the entire length of coax is protected from pickup and reradiation as well as outer surface conduction.

To test the results of the modified coax, I placed a 5-inch antenna probe at the operating position a few inches from the equipment front panels. I connected the probe to my spectrum analyzer and compared the modified coax with an unmodified coax. The reduction in RF pickup while transmitting from 80 meters to 10 meters was between 12 and 18 dB—a very significant improvement that, by the way. further reduces operator exposure to RF fields. The 15-foot RG-58 coax, when terminated with a 50-ohm load, measured about 0.2 dB loss of the desired signal at 29 MHz.

Antenna impedance

I own a General Radio Model 1606A RF Impedance Bridge (vintage 1950s) that's excel-

lent for HF antenna measurements. These bridges are sometimes available from surplus outlets. I used the setup in **Figure 4** to make 113 impedance measurements, each requiring about 5 minutes. I applied several correction factors to the raw data, using a computer and a spreadsheet program.

Figure 5 shows plots of the series resistance and reactance values. Note the following:

A) The system is series resonant at approximately 1.8 MHz.

B) A high-Q parallel resonance occurs at 3.4 MHz. The typical resistance peak occurs and the reactance goes through a large inductive to capacitive transition.

C) At higher frequencies, the reactance goes through periodic series and parallel resonances; but, the maximum to minimum variations decrease greatly. The peaks are separated by the half wave frequency of the wire—about 3.4 MHz, as are the minimum points.

D) The series reactance tends to move in a capacitive direction at higher frequencies. I couldn't find any measurement errors that might explain this. The resistance also drifts toward lower average values at the high end. These results may be due to the increasing height, in wavelengths, and the effects of imperfect ground underneath the antenna.[1,2]

E) The resistance excursions also become smaller. Toward the high frequency end the "loops" appear to be larger, but this is an optical illusion caused by the logarithmic vertical scale.

F) The minimum resistance value is always greater than 50 ohms, above 2.5 MHz. Looking at the Smith Chart in **Figure 6**, you'll see this means that all R and X values are within the 50-ohm circle. A further consequence of the high antenna resistance is that the resistance of the RF ground system is less degrading to the antenna efficiency.
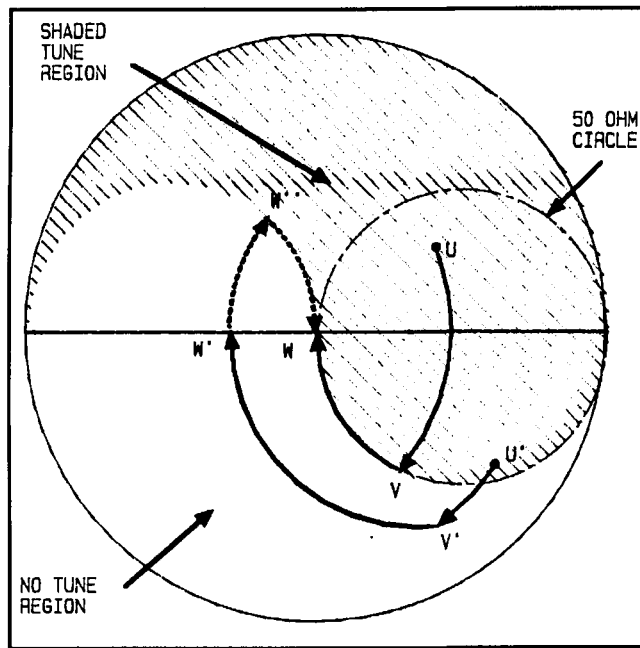
Figure 6. Smith chart used to illustrate methods of tuning the L network to 50 ohms.

G) Some aberrations in several of the data point values may be due to measurement errors or bends in the wire (**Figure 1**). The 1606A bridge is a bit tedious to operate when making antenna measurements at the higher frequencies, due to hand proximity effects.

H) Because of the periodicity of the long-wire antenna, it is an efficient radiator of harmonics, therefore the transmatch should be designed to give plenty of help in that respect.

The transmatch design

One advantage to having a set of accurately measured driving point impedances values for
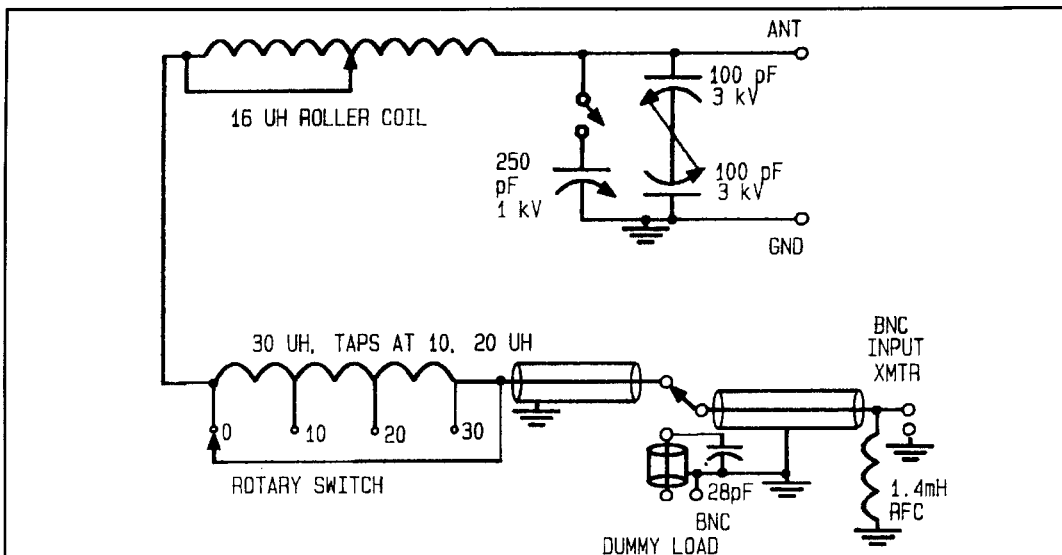
Figure 7. Schematic of long-wire antenna coupler.

Photo A. Long-wire antenna coupler.

war transmitter. The roller coil and the two tuning capacitors have logging scales, so you can use a tuning lookup table. The SWR indicator is a separate item.

Referring to the Smith Chart of **Figure 6**, the susceptance of shunt C moves impedance point U down to the 50-ohm circle at point V and the series L tunes out the reactance to the 50-ohm pure resistance point at point W. There is one problem, though. If the impedance point is at U', then a small value of shunt capacitive susceptance will take us outside the 50-ohm circle at V' and the series coil can only get us to point W', which is unacceptable. Therefore the minimum value of shunt C must be small enough, as required by the worst case antenna impedance values. I solved this by using the circuit arrangement in **Figure 7** and by mounting the roller coil, the tapped inductor, and the inductor switch on Plexiglas™ blocks and doing whatever else I could to reduce stray C. As an additional measure, you could increase inductive reactance to point W" and use a shunt capacitor at the transmatch input to get to point W. This is the Pi network, of course, which I wanted to avoid because it detracts from the simplicity I preferred for this system. On the other hand, when I need a larger value of C at the output end, I connect the 250-pF variable with a knife switch. With the values shown, I can tune my antenna easily and continuously from 2.5 to 30 MHz. Below 2.5 MHz, for receiving only, I use other approaches that are best for the specific frequency ranges of interest.

RFC in **Figure 7** is a heavy-duty transmitting choke that displays no series resonances or

the antenna is that you can determine the transmatch L and C values at any single frequency using a Smith Chart (**Figure 6**) or lumped-circuit analysis, and a circuit design can be established that emphasizes harmonic reduction. For the impedance values that I measured, all inside the 50-ohm circle, the L network of **Figure 7**, with inductance in series followed by shunt capacitance, worked out quite well. **Photo A** shows the unit, which uses materials either on hand or purchased from surplus stores. The two variable capacitors and the large dial were purchased new in 1946 and used in my first post-

Figure 8. Method for calculating L and C. See text and Appendix B.
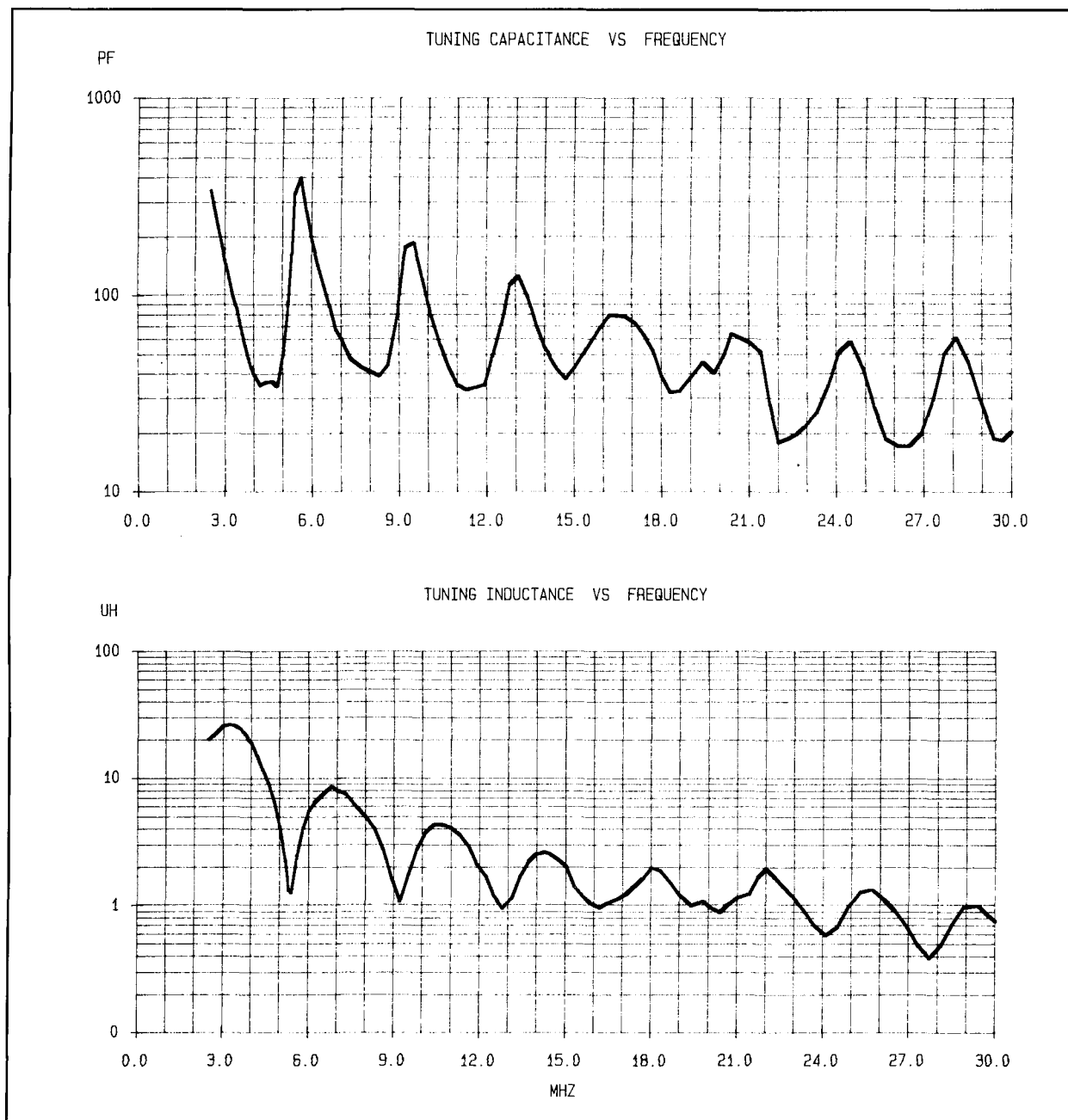
Figure 9. Calculated values of inductance and capacitance for the transmatch.

"suckouts" in the 2 to 30 MHz range and doesn't affect tuning. It acts as a backup to the transmitter's internal circuitry; in the event of a transmitter parts failure, it keeps dangerous voltages off of the antenna wire and the transmatch components. Better safe than sorry! It also drains off static charge buildup. Be sure to short out the unused segments of the two inductors. I had all kinds of problems with spurious coil resonances that made several frequencies impossible to tune. If the shorted turns and the shorting connection were perfect conductors, no power—coupled from the active segment of the coil—would be lost. In practice, some small

amount can be dissipated. It's best to accept this fact.

The 28-pF capacitor, two 56-pF 500-volt dipped micas in series, is necessary because the ceramic wafer switch puts a "bump" in the coax line impedance. This capacitor was test selected to improve the return loss into the dummy load from 20 to 40 dB across the entire DC to 30-MHz band. This makes the dummy load an almost pure 50 ohms resistive.

Transmatch component values

Knowing the transmatch circuit, it's possible

CAPACITOR VOLTS VS FREQUENCY FOR 100 WATTS OUTPUT
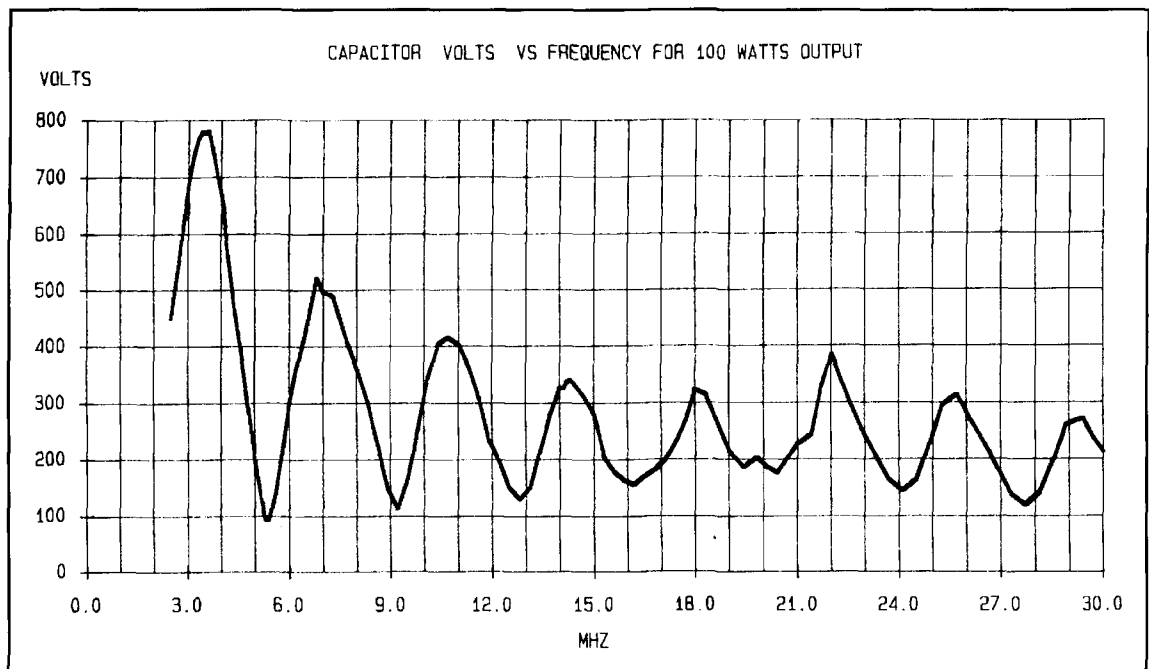
Figure 10. Calculated values of capacitor voltage.

to determine everything important about the transmatch values and component specifications. **Figure 8** shows the basic idea. At a specific frequency, the antenna RA and jXA are known by measurement. The L and C values for each frequency are then calculated by a computer spreadsheet program and the driving point impedance ZIN is 50 + j0 ohms (i.e., S11 = 0). At this point, the input impedance of the transmatch is the same as the dummy load used to tune the transmitter, therefore no further transmitter tuning is needed and the forward power into the transmatch is the same as into the 50-ohm dummy load. This may be a valid reason for tuning fairly close to 1:1 VSWR.

A possible complication with regard to the statement above occurs if the transmitter output contains a significant amount of available harmonic power. The transmatch reflects these harmonics, and a small amount of VSWR may be indicated due to the harmonic return waves. The dummy load wouldn't reflect these harmonics. Assume that this effect is below the

threshold of the SWR meter because of preliminary harmonic filtering within the transmitter.

After determining L and C values, you can calculate the voltage on the capacitor, the antenna current, the power loss in the inductor and the ground connection, and the harmonic reduction. Some of the mathematical methods used are in **Appendix B**.

Figure 9 is a plot of calculated L and C values. They show the same kind of periodic variations as in **Figure 5**. The large peak in C value and dip in L value at 5.5 MHz occurs because the antenna's series resonant impedance (**Figure 5**) is close to 50 ohms resistive and low reactance. This is a "don't care" frequency, as far as transmitting is concerned. If a situation like this occurs inside a ham band, changing the antenna length a few feet will move it out of the way. With the L and C components I used, I could tune pretty well at 5.5 MHz. At frequencies used only for receiving, the exact match is usually not critical.

Figure 10 is a calculated plot of capacitor
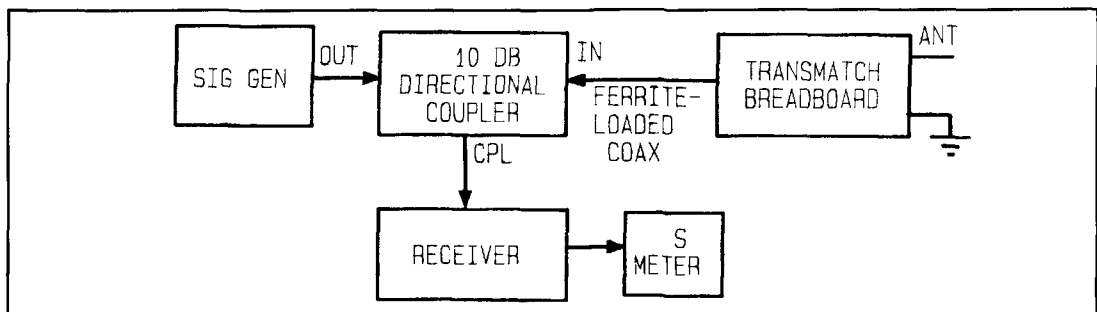
Figure 11. Setup for determining L and C values using a directional coupler, such as a Mini-Circuits type ZMDC-10-1.
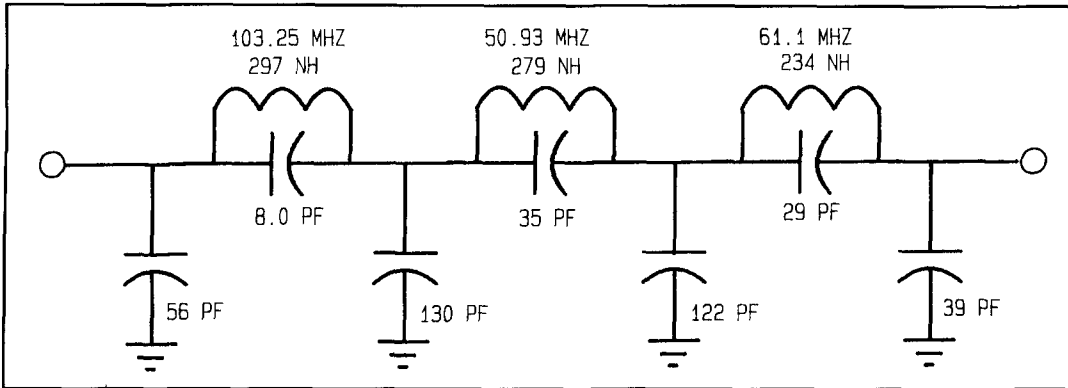
Figure 12. Low-pass filter schematic.

voltage for an output power level of 100 watts. It becomes fairly large on 80 meters, but is much lower on other ham bands. If I had increased the length of the wire about 6 feet or so, I would have avoided this problem on 80 meters. I used capacitors of 1000 volts (C1) and 3000 volts (C2) because they were on hand.

I describe an approximate method for estimating harmonic reduction, with some reservations. Apply 1.0 volt from a 50-ohm source at the fundamental frequency and calculate the power in the resistive part of the antenna impedance at the fundamental frequency.

Double the frequency of the 1.0-volt generator and use the values of antenna resistance and reactance for the second harmonic frequency, then calculate the power in the resistive component. I calculated the following example values (using a SPICE simulation program) of second harmonic reduction, assuming ideal transmatch components and circuitry:

3.5 MHz	26 dB
7.0 MHz	38 dB
14.0 MHz	63 dB

The main error source with this method is
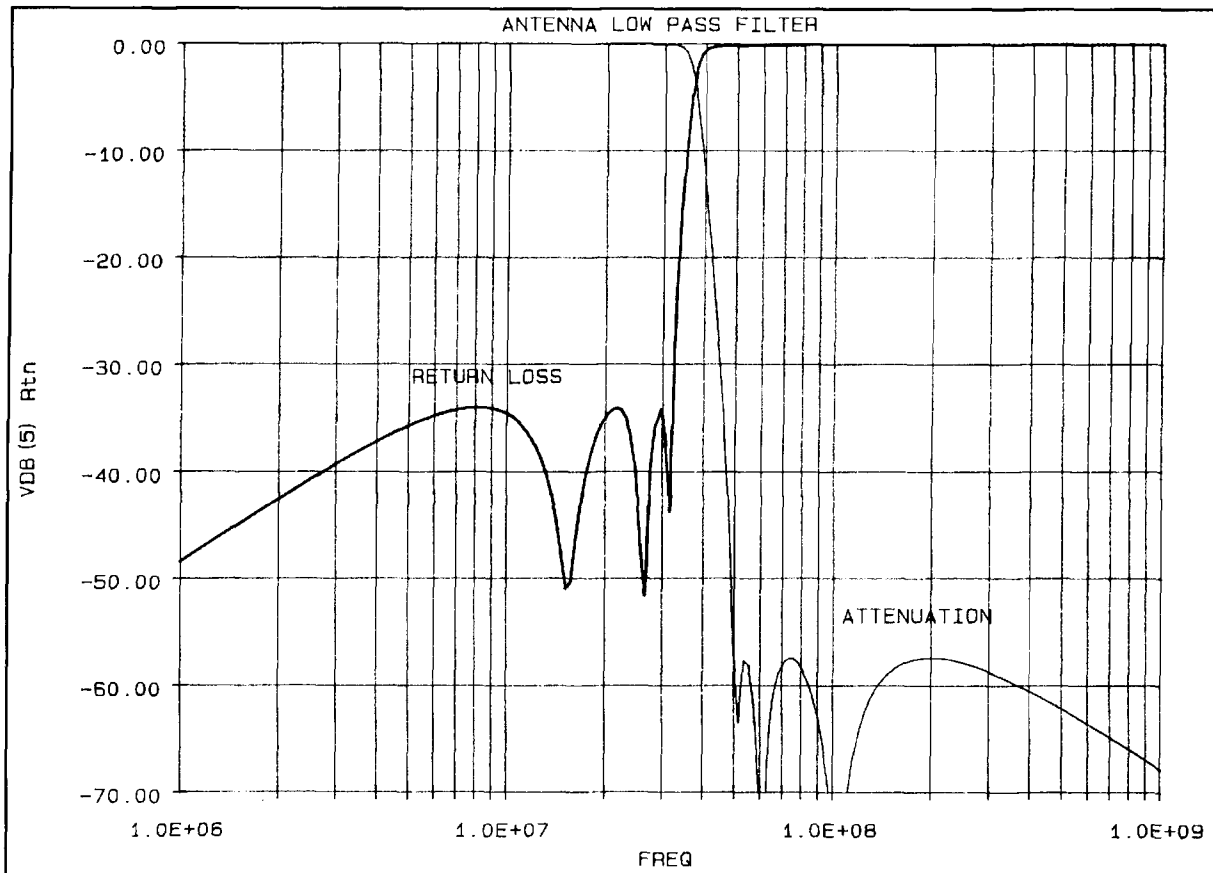
Figure 13. Low-pass filter response.

that the transmitter output impedance, as transformed by the length of the 50-ohm coax, can be much different than 50 ohms resistive,[3] and this affects the low-pass properties (but not the input impedance) of the transmatch. For example, if the output impedance of the coax is very large (a current source), the coil is less effective in reducing harmonics. If the output impedance is very small, the coil will have greater influence. Solid-state PAs tend to have low output impedance; but the length of coax to the transmatch could transform this value considerably. Vacuum tube PAs can have a wide range of output impedance values. The 50-ohm test is the best attempt at estimating harmonic reduction, without further information.

There is one potential problem with this transmatch and a long-wire antenna. Using my scalar network analyzer, scanning from 2 to 30 MHz, and looking at return loss on the coax, at almost any tuneup setting I could see certain other frequencies, both higher and lower, where attempts at impedance matching were occurring. The antenna impedance and LC values have the right combination to produce at least some small degree of match at these frequencies. Furthermore, these frequencies are separated by approximately the half wave frequency. An examination of **Figure 9** shows similar L and C values for frequencies separated in this manner. This explains why the second harmonic attenuation of 3.5 MHz, calculated above, isn't larger. If the wire were a few feet longer, I'm sure this number would improve.

If the transmitter happened to have a spurious output at one of these frequencies, you might experience some interference. When using a long wire antenna with a low Q transmatch such as this one, verify transmitter purity across a wide frequency range. In a collocated environment where transmitters at these special frequencies may be present, some receiver types may be a little more vulnerable.

I also evaluated a borrowed transmatch of the SPC (Series-Parallel-Capacitor) type.[4] Because of its much sharper bandpass selectivity, this transmatch was free of the superfluous responses mentioned above. It requires frequent razor sharp retuning, especially on the lower frequency ham bands. Still, I stayed with the L network because I had no problems with transmitter nonharmonic outputs, as verified on my spectrum analyzer. I didn't experience any difficulties while preparing a tuning table. The settings for 1:1 VSWR were unique and unambiguous. My L network is simple and easy to build from easy to obtain parts, it's very easy to tune and does a good job if you choose the antenna length properly, as discussed before.

An alternative approach

In the previous discussion, I used antenna impedance measurements obtained from special test equipment. **Figure 11** shows another method for obtaining measurements. By employing this setup plus a hand-held digital capacitance meter and a grid dip meter to determine inductance values by resonating with a known value of C, it's quite feasible to find the right values of L and C for the ham bands using small receiving-type components like trimmer capacitors and slug-tuned coils. Most amateurs can find values of adjustable C between 15 pF and several hundred pF and values of adjustable L between 0.2 µH and 50 µH can usually be found in their parts collections. Tune the network for minimum signal into the receiver. Locate the experimental setup at the position where the final transmatch will be located. Once you know the values of L and C, you can use a Smith Chart to find the antenna impedance values, if desired. Also, once you know the range of values, it's just a matter of collecting the right transmitting components. The final transmatch should have very low stray C and L in the layout.

Low-pass filter

To enhance the reduction of VHF/UHF harmonics, I mounted a low-pass filter, shown schematically in **Figure 12**, adjacent to the transmatch input (see **Figure 1**). This seven-section elliptic filter is listed in the catalog[5] as CC0702, $\theta = 40$ degrees. This filter has a VSWR of 1.04 or better (35 dB return loss) below 30 MHz and at least 55 dB attenuation above 50 MHz, as shown in **Figure 13**. The measured response, with 50-ohm source and load, was within a few dB of theoretical. As mentioned before, the in-system VHF/UHF response can be modified by the source and load impedances of the system, which at VHF/UHF are seldom 50 ohms and can vary unpredictably. The compromise procedure is to pick the textbook capacitor values with a digital C meter and then adjust the coils to resonate at the indicated frequencies using a grid dip meter, frequency counter combination. This optimizes the filter passband input VSWR (if the transmatch is tuned to 50 ohms). We then accept a certain amount of stopband inaccuracies.

Additional thoughts

There are three main motivations for this project. One is the learning experience of designing, understanding, and building the complete antenna system, including the transmatch. Another is to optimize the performance of the antenna despite its simplicity and handicaps. The third is to obtain adequate amounts of har-

monic reduction with a very simple circuit, which is desirable in a long-wire antenna. The antenna wire itself should have a length that will tune nicely in the ham bands. In a long-wire antenna system, if resistive losses and ground losses are minimal, there's no reason for the wire length to be exactly resonant at some particular frequency—as long as it can be tuned easily. In fact, if it's a half wavelength on the lowest ham band, the RF voltage at the transmatch becomes quite large, as happened in my case. It's preferable to keep maximum voltage out in the back yard somewhere, and maximum current up as high as possible. Also try to avoid lengths that provide maximum current in the RF ground connection resistance, in order to improve efficiency. The antenna simulation program is an excellent tool for prelocating voltage and current maxima and minima.

If the circuit is modified so the capacitor can be switched back and forth between input and output, this transmatch will also transform any point inside the Smith Chart to 50 ohms. Some end-fed wire antennas may require this at their lower frequencies.

Acknowledgment

I appreciate several valuable suggestions by Brian Beezley, K6STI, which have been taken into consideration in this article.

Appendix A

I used the ELNEC program[6] (based on MININEC[7]) to get a "feel" for the behavior of my long-wire antenna at various angles of elevation. Because these programs don't model "real" ground directly underneath the antenna, I couldn't obtain close agreement with measured impedance values or dependable values of radiation strength at high angles of elevation for an antenna of such low height.

The aluminum siding on my house was modeled as a gridwork of noninterconnected horizontal and vertical "wires," each 10 inches in diameter. The effects of the structure in attenuating the signal toward the house and in reflecting toward the far end of the antenna were evident (but less pronounced than I expected) when I compared results with and without the presence of this gridwork.

The effects of the sloping vertical component of the wire were also apparent. At frequencies where the current in the vertical portion is significant, it contributes a considerable amount of vertically polarized radiation—especially off the ends of the horizontal antenna axis. This helps to improve the general coverage properties, so the radiation from this segment isn't really wasted. If the height of the flat top segment can be increased, the length and effectiveness of the vertical segment can be improved, especially at the lower frequencies. Simulations also show that putting another sloping segment at the far end not only enhances low angle radiation at the higher frequencies, but also extends the low frequency limit.

In general, the patterns for my installation (**Figure 1**) agree pretty well with handbook patterns. The patterns at 25 to 30 degrees elevation are strong enough to enable some DX work, with fairly deep nulls in various directions that are improved somewhat by the vertical component, which has been amply verified in practice (but of course things could be a lot better). The patterns at 35 to 50 degrees are stronger and have shallower nulls for E-layer work in domestic contests and general operating. For the DX-minded operator, an additional 10 feet of antenna height would be helpful.

Appendix B

From **Figure 8**:

$$G_A + jB_A = \frac{R_A}{\left(R_A^2 + X_A^2\right)} - j\frac{X_A}{\left(R_A^2 + X_A^2\right)} \qquad (1)$$

$$G + jB = G_A + j\left(\omega C + B_A\right) ; \ G =$$

$$G_A \text{ and } B = \omega C + B_A \qquad (2)$$

For any point on the 50-ohm circle of a 50-ohm Smith Chart:

$$0.02 = \frac{G^2 + B^2}{G} = \frac{G_A^2 + B^2}{G_A} \qquad (3)$$

Solve **Equation 3** for B (use the positive root so B is a positive [capacitive] susceptance), insert the result into **Equation 2** and get:

$$G + j\sqrt{0.02\ G_A - G_A^2} = G_A + j\left(\omega C + B_A\right) ;$$

$$\text{if } 0.02\ G_A \geq G_A^2 \qquad (4)$$

Equate the imaginary parts of **Equation 4** and solve for the value of C that translates the antenna impedance to the lower periphery of the 50-ohm circle:

$$C = \frac{\sqrt{0.02\ G_A - G_A^2} - B_A}{\omega} ; \ \sqrt{0.02\ G_A - G_A^2} > B_A \qquad (5)$$

Next, also from **Figure 8**:

$$R + jX = j\omega L + \frac{1}{G + jB} = j\omega L + \frac{1}{G_A + jB} \quad (6)$$

Set the imaginary part of the right side equal to zero and solve for the value of L that gives a 50-ohm pure resistance input impedance:

$$L = \frac{B}{\omega\left(G_A^2 + B^2\right)} = \frac{\omega C + B_A}{0.02\ \omega G_A} \quad (7)$$

To find the capacitor voltage at 100 watts output:

$$V_{IN} = \sqrt{(100)\ (50)} = 70.7\ \text{V} \quad (8)$$

$$|V_C| = 70.7\sqrt{1 + (\omega L\ /\ 50)^2} \quad (9)$$

REFERENCES

1. C. Michaels, "Dipoles Above Real Earth," *QST*, November 1992, page 67.
2. J.D. Kraus and K.R. Carver, *Electromagnetics*, 3rd edition, McGraw-Hill, New York, 1973, Section 10–14.
3. W.B. Bruene, "RF Power Amplifiers and the Conjugate Match," *QST*, November 1991, page. See also "Technical Correspondence," *QST*, May 1992, page 95.
4. *The ARRL Antenna Book*, 1991 edition, pages 25-8 to 25-11.
5. A.I. Zverev, *Handbook of Filter Synthesis*, John Wiley and Sons, New York, 1967.
6. ELNEC, a personal computer antenna simulation program by Roy Lewallen, W7EL, 5470 SW 152 Avenue, Beaverton, Oregon 97007.
7. AO 5.0, a MININEC-based Antenna Optimizer by Brian Beezley, K6STI, is also an excellent program.

PRODUCT INFORMATION

Harris Introduces Multi-Waveform Serial HF Radio Modem

Harris RF Communications has unveiled a new serial modem providing data communications up to 4800 bps.

The RF-5710 is for use in naval broadcast and ship-shore data communications, long-range point-to-point data communications, secure digital fax, narrow-band secure voice and digital video image transmission (DVIT).
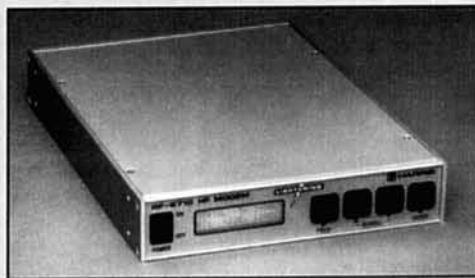
The RF-5710 uses a patented data directed equalizer to eliminate the effects of intersymbol interference due to multipath. The modem tolerates large multipath spreads and fast fade rates. The modem also includes convolutional coding, bit interleaving and Viterbi soft decision decoding.

The waveform, data rate and other modem parameters are selectable using the front-panel keys and LCD display, or by way of a remote control interface.

For more information, contact Harris Corporation, RF Communications Group, 1680 University Avenue, Rochester, NY 14610; or phone 716-244-5830.

Larsen switches to polycarbonate/PET (polyethylene terephthalate)

Larsen Electronics, Inc., manufacturer of Larsen® Kulrod® mobile antennas for citizens band, amateur, commercial radio, and cellular phones, was having a problem with antenna bases cracking. To increase the base's chemical resistance, the company identified Makroblend polycarbonate/PET resin, manufactured by the Polymers Division of Miles Inc., Pittsburgh, as a possible substitute. The Makroblend UT-1018 resin resisted chemicals and ultraviolet tests. Larsen is now using the resin on a wide variety of their Kulrod antennas.

For more information about Makroblend UT-1018 polycarbonate/PET resin, write: Polymers Marketing Communications, Miles Inc., Mobay Road, Pittsburgh, PA 15205-9741 or phone: 412-777-2000.

For more information about Larsen Kulrod mobile antenna bases, write: Larsen Electronics, Inc., 3611 N.E. 112th Avenue, P.O. Box 1799, Vancouver, WA 98668 or phone: 206-944-7551.

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

DUMMY ANTENNAS

Eliminate on-the-air tune-up sessions

A dummy antenna is intended to behave just like a real antenna in all respects, but one. It does not radiate. Any power fed to a dummy is dissipated as heat in a resistor. Also, any incidental fields the dummy antenna produces are blocked by shielding, so no power escapes. In the same way, it does not pick up external signals.

The dummy loads we use for transmitter testing are not dummy antennas, because they are designed to behave as a matched section of transmission line. Most are for "50 ohm" line, and are a 50-ohm low inductance power resistor. Some may use the more exact value, 51.5 ohms. Other values include: 75, 290, 350, 400, and 600 ohms—the first for coax and the others for open wire lines. Since these have no reactive component, they cannot simulate an antenna—except for one that has resistance of 50 ohms at resonance (or any of the other design values). Proper dummy loads do not radiate, although it is sometimes instructive to reply to a CQ using a dummy load instead of an antenna.

The difference between dummy loads and dummy antennas is illustrated by the resistance and reactance curve of the antenna in use. **Figure 1** shows the resistance and reactance for two dipoles as a function of frequency, from very low frequency to past the fourth or full-wave resonance value. The two curves represent a thin wire antenna and one with a large conductor diameter. The pattern of change repeats at each additional resonance point, with some difference in the R, X values.

The duplication of such impedance variations over any desired frequency range in a small non-radiating package is theoretically possible, but requires at least one LCR combination for each resonance point. Such a network is no great problem at the low power levels of receivers, but is physically difficult at transmitter power levels. For this reason, dummy antennas are usually designed to cover only a part of the possible frequency range of an antenna. A typical simple design may be useful at the low frequency end, or at a point close to the first resonance. Slightly more complex designs may simulate the antenna from a low frequency up to first resonance, or a bit beyond. These cover most needs, and are the only type considered here.

Historical dummy antennas

A look in most of the older *Radio Engineering* handbooks or textbooks will show some dummy antenna designs that have been used in the past. **Figure 2** presents two of these, the one at A specified by the IRE (1948) for broadcast and shortwave receivers, and that at B for auto receivers.[1] Both are intended to be placed in series with the line from the signal generator used to the receiver.

In that period, most receivers were designed for a simple L antenna operating against ground. The dummy reflects this use. At low frequencies, around 600 kHz, the dummy is essentially the 200-pF capacitor, with a reactance around 1000 ohms. At frequencies above 4 MHz, the dummy is essentially the 400-ohm resistor. Minimum series impedance occurs at 2 MHz, where the dummy resistance and reactance are equal to about 150 ohms each.

The auto antenna is intended to represent a short whip antenna, and is essentially a resistor in series with a leaky capacitor. The capacitor is larger than the capacitance of the antenna alone, since the normal auto installation uses a length of "low capacitance" shielded cable to connect the antenna to the receiver. This has very little effect as an impedance transformer, but does add capacitance plus a little loss across the antenna base. Other dummy antenna designs that have been used are 100 pF in series with 20 ohms, and merely a 40-pF capacitor.

The broadcast type dummy is useless with coax input equipment, or with ferrite loops. The auto type is still useful, but rarely found.
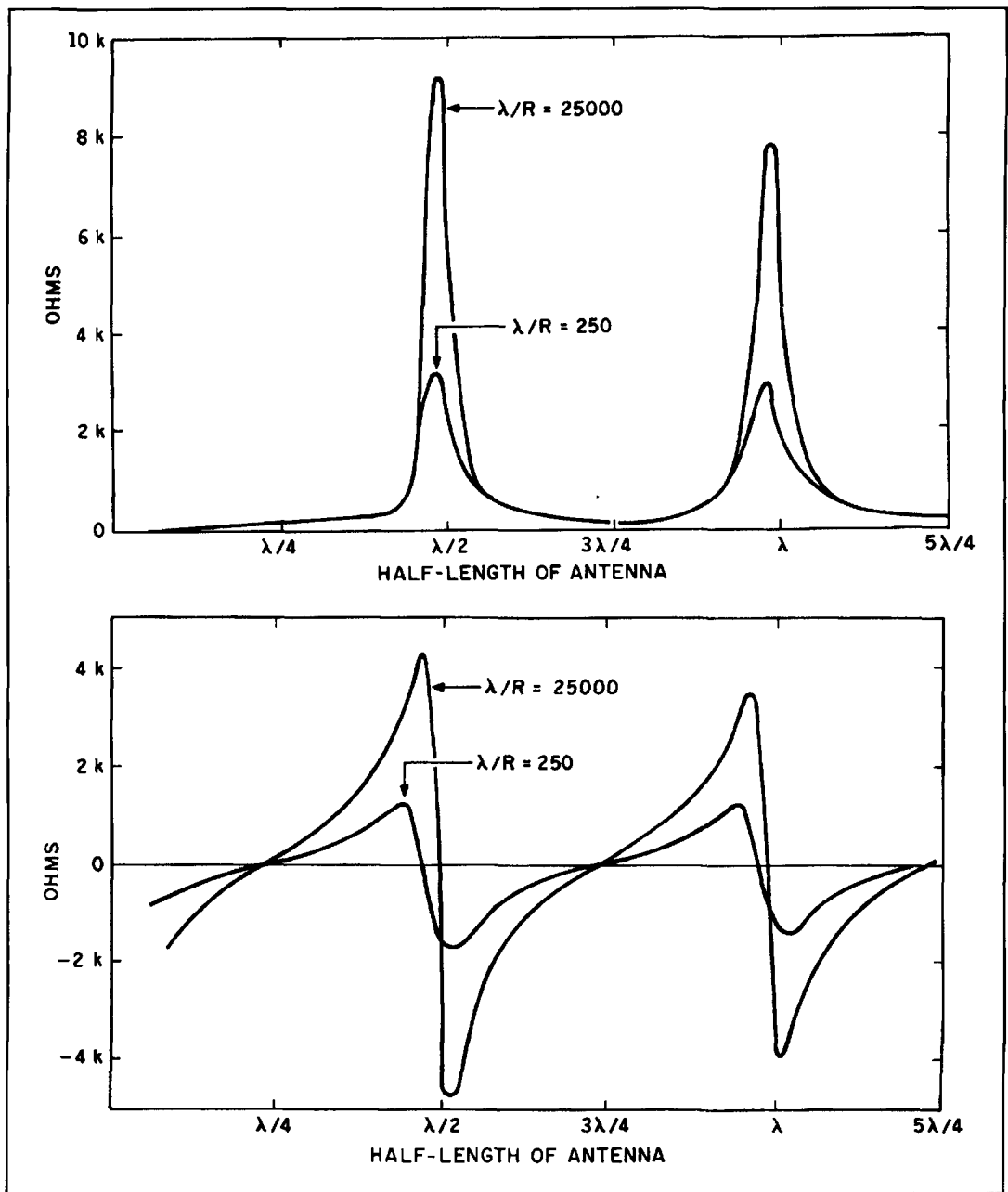
Figure 1. Dipole resistance and reactance. Impedance variation of two dipoles of the same length, but different diameter. (A) is the feedpoint resistance, and (B) the feedpoint reactance. Note the insensitivity of feed resistance to diameter at the quarter wave resonance points, and the shift in resonant frequency with diameter. The patterns repeat each additional half wave, with both resistance and reactance reaching a constant value..

Today, almost all of the auto receivers are installed at the factory, and there is little repair required. Consequently, dummy antennas are rarely found today—even in well equipped stations or labs. However, they still have a useful place in design, experimentation, and even in normal station operation. Some designs and use notes follow.

Small dummy antennas are occasionally found on hamfest swap tables. The most common are World War II surplus types, intended for tune-up of the ARC series of airborne trans-

mitters while the aircraft is sitting on the ground. They are of the simple type—a resistor and capacitor in series—simulating an aircraft trailing a wire antenna.

Single band dummy antennas

In most antenna systems, the impedance at the feedpoint for the band of interest is defined by the linear relations shown in **Figure 3**. The reactance component is zero at or near the point

of minimum SWR, and changes linearly with frequency change, or nearly so. The resistance component increases with frequency, again nearly linearly. These curves are for a dipole. Multiple element antennas may show more complex variation, especially if complex matching elements are used. (See Maxwell,[2] for more examples of measured values for dipoles and a beam.)

A first-order dummy antenna for antennas with this pattern of impedance variation requires only three components in series: Ld, Cd, and Rd, as shown in **Figure 4**. Ld and Cd must resonate at the antenna resonant frequency, and be sized to give the same reactance variation as the antenna as frequency changes. Rd is the antenna resistance, less the RF resistance of the coil used. This simple approach to resistance is exact at only one selected frequency, usually chosen to be the resonant frequency. The simulation error will increase for frequencies away from resonance, but will not be serious across the design amateur band

To design a dummy antenna (system) to match your own antenna system, you need two or preferably three or more measurements of R and X at the antenna feedpoint. An RF bridge is the ideal measurement method, but low cost "antenna bridges" will provide useful data. Correct the measured values for any transmission lines used (Maxwell,[2]). Plot the measured points on a graph, and draw the best fit straight R and X lines, as in **Figure 3**. Choose a value of resistance that is the average of the measured resistance at the band limits, or the frequency range of interest. This is Ra. Note the frequency of resonance, Fr, in MHz (zero reactance), pick a point on the reactance line, say at the band edge, and note the frequency F, MHz, and reactance X ohms, which may be positive or negative. Let B equal F/Fr. Then, from a relation in Terman,[4] the coil inductance is, in computer notation:

$$XL = X/(B/(B^2-1)) \text{ ohms}$$

$$L = XL/(2*PI*Fr) \text{ microhenries}$$

The value of C is:

$$XC = -XL \text{ ohms}$$

$$C = 1/(2*PI*XC) \text{ pF}$$

Based on coil construction, measure or at least estimate the coil Q. For a small coil for a receive-only dummy this may be around 100, and for a large coil for transmission, around 250. The coil resistance is:

$$Rc = XL/Q$$

and the value of the series resistor is:

$$R = Ra-Rc$$
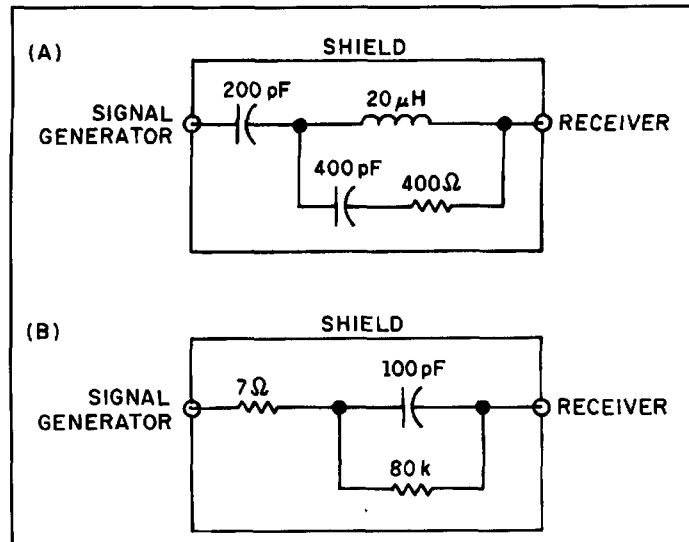
Figure 2. Historical dummy antennas. (A) IRE standard dummy antenna for radio receivers, 1948. (B) Simpler dummy used to simulate short auto antenna whips. Both are intended to be used in series with the signal generator to receiver connection.

For example, for the dipole of **Figure 3**, L = 35 μH and C = 50 pF. The average Rd is 64 ohms. For a Q of 250, R = 60 ohms (values rounded to available component sizes).

For a receive-only dummy, these components may be a small air or slug core coil, a silver-mica capacitor, and a 1-watt carbon resistor, or combinations to obtain the exact value. For transmitting, they must withstand the power used. For a 100-watt transmitter, the current will be about 2 amperes. Number 16 wire is ample for the coil. The capacitor may be a transmitting ceramic, or several of the large (about 1-inch square) mica capacitors in parallel. Mount the assembly in a shield box with a connector as used in the station.

If you are developing equipment, an assortment of these dummy antenna boxes for the bands you are using is invaluable.

Simulating the antenna-transmission line system

These simple dummy antennas are fine, but they simulate the impedance at the antenna. For use in tuning up the transmitter and tuning unit, the simulation must be for the impedance at the feed end of the transmission line. In other words, what is needed is a "black box" that will simulate the combined effects of the antenna and transmission line.

Any of the reference books on RF circuit theory will show the technique of constructing a network to simulate a known impedance. While this is not really difficult, there is a drawback. If either the antenna or transmission
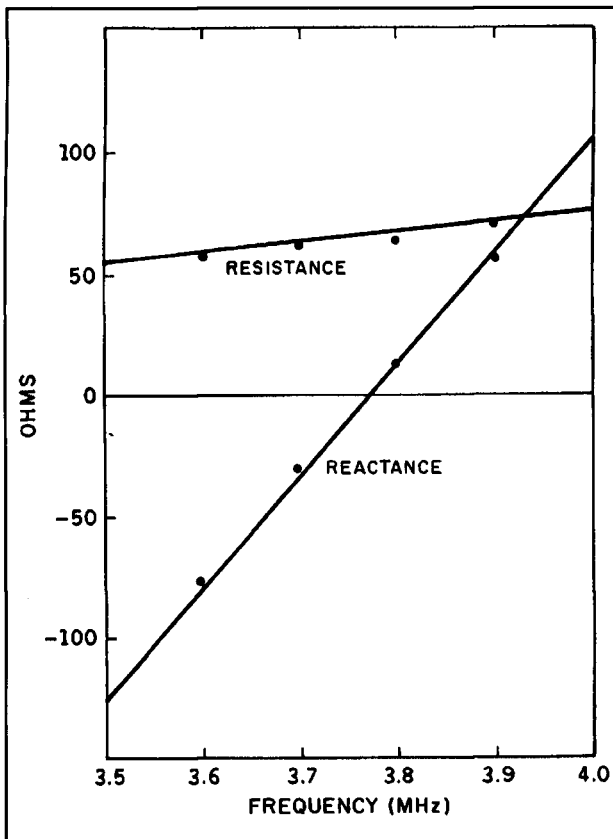
Figure 3. Measured dipole impedance. Measured values of an 80-meter dipole at 40 feet elevation, across the 80-meter band (data from Maxwell, 1989). If the antenna being measured has high SWR at the band edges, the reactance curve will develop curvature, noticeable at the band edges.
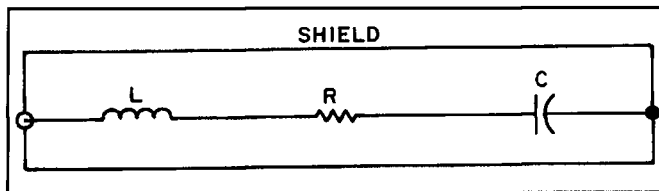
Figure 4. Narrowband dummy antenna. Simple CRL network to simulate a dipole or monopole antenna across an amateur band. Simulation error can be ideal at only the design frequency, usually at or near band center.

line is changed, the black box is no longer correct, and the sequence of measure, design, and construct must be repeated. Separate simulation of antenna and transmission line reduces the work involved.

As shown in **Figure 5A**, the simplest simulation of a transmission line is another line of the same characteristics and length. For receiver or low transmitter power, the line can be RG-174/U. This line is small enough to fit into a small shielded box, even though fifty or a hundred feet is needed to equal the station transmission line. The box connects to a low power type of dummy antenna. For high power, the simulator should be the same type and length as

the regular transmission line. The difference is that this unit terminates in a power type of dummy antenna.

There is a way to avoid the extra transmission line if only a single-band antenna is connected to the line. This is to make the regular transmission line a multiple of a half-wave in length at the center frequency of the band. Make the measurements of R, X at the feed end of the transmission line. Use these in the above equations with no correction for line length.

To use, switch to the dummy antenna (plus the "dummy line" if you've included one), tune up, then switch to the regular antenna and start operating. If the reflected power at the transmitter input is appreciable, touch up the antenna coupler tuning to bring it to zero. Do the same if you QSY. The benefit? No QRM to other stations, and good performance for you.

The experimentally inclined might like to try a true artificial line of the type shown in **Figure 5B**. These are a series of Pl sections, each designed for a different frequency. (See Henney,[1] for design relations.)

Wide band dummy antennas

A dummy antenna of the type shown in **Figure 4** can be used at higher or lower frequencies in addition to the band for which it is designed. However, the matching error becomes very large. More components are needed to improve the match simulation.

Of the designs in the literature, a four-component dummy developed by computer simulation seems to be best for amateur use .[3] It is intended to provide a very close match to a dipole for frequencies below resonance, and a reasonably good match through resonance and up to about 1.2 times resonance. This covers a lot of antenna situations—including the mobile whips, verticals, and dipoles used on lower bands.

The connection diagram is shown in **Figure 5**. In essence, the series R and L of **Figure 4** has been transformed to R and L in parallel, then bridged with a small capacitor. All values have been adjusted for good overall simulation.

Calculation for design values are based on a cylindrical dipole antenna of radius a meters. The distance from center to end (half height) is h meters. The design equations are, again in computer notation:

$$C31 = 12.0674*h/(LOG(2*h/a)-0.7245) \text{ pF}$$

$$C32 = 2*h*((0.89075/((LOG(2*h/a)^0.8006)-0.861)-0.02541)) \text{ pF}$$

$$L31 = 0.2*h*(((1.4813*LOG(2*h/a))^1.012)-0.6188) \text{ μH}$$
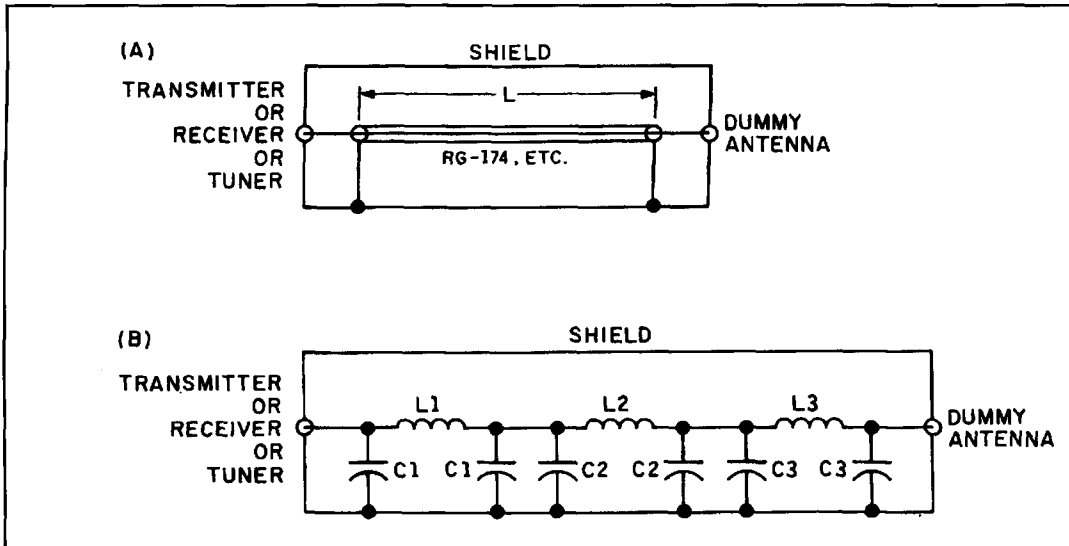
Figure 5. Artificial transmission lines. Simple and complex emulation of a transmission-line dummy antenna-to-transmitter connection, to present the same conditions as the real systems. (A) shows the simple system, a small diameter line of the same impedance and length as the regular one. (B) shows emulation by a series of PI sections. The accuracy of the simulation increases as the number of sections increases.

$R31 = (0.41288*(LOG(2*h/a))^2)+(7.40754* LOG(2*h/a)^{-0.02389})-7.27408$ kilohms

For a monopole (vertical), use h as the overall height, divide the L and R values by two, and multiply the capacitor values by two.

For checking, the data and calculated impedance of the dummy antenna of **Figure 5** are:

h   = 0.9 meters
a   = 0.00264 meters
h/a = 341

C31 = 5.149 pF
L31 = 0.6537 µH
R31 = 2,380 k
C32 = 1.067 pF

Length 2h Wavelengths	Frequency MHz	Z, R+jX Ohms
0.125	20.83	3.15 + j1397
0.250	41.67	13.50 + j563
0.373	62.5	34.27 + j218
0.500	83.33	73.13j + 39.29

The simulation error compared to a dipole is about 10 percent at the half-wave point, less at lower frequencies (see **Figure 6**).

If you are working on phased verticals to take advantage of the lower sunspot numbers, a set of these wideband dummies can be a big help in setting up the phase shifting networks. Use a good oscilloscope for measurement of phase and amplitude.

Concluding comments

It is easy to see why dummy antennas have fallen out of favor. The need for a dummy for
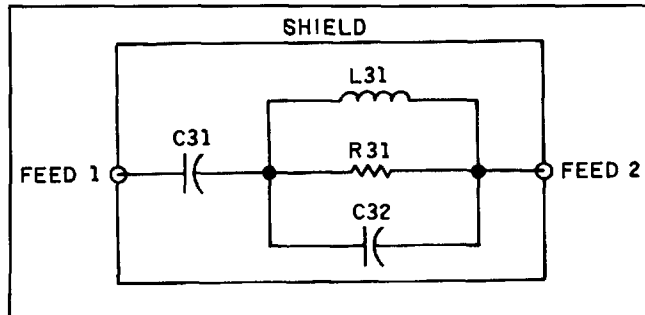
Figure 6. Wideband dummy dipole antenna. Dummy antenna to emulate the impedance of a dipole from very low frequencies to about 1.2 times its resonant frequency. For coax, feed 2 would be grounded.

each amateur band as compared to the convenience of a 50-ohm dummy load is probably the major factor. Then too, the current lack of interest in home construction is a factor. And note that manufacturers of antennas never offer a dummy antenna version to aid in tuneup of the system.

Despite these factors, the dummy antenna merits more use. The major gain can be the elimination of those tune-up sessions heard on all bands.

REFERENCES

1. Keith Henney, editor, Radio Engineering Handbook, McGraw-Hill, 1959.
2. M. Walter Maxwell, Reflections, Transmission Lines and Antennas, American Radio Relay League, Newington, Connecticut, 1960.
3. Tee G. Tang, Quang M. Tieng, and Morris W. Gunn, "Equivalent Circuit of a Dipole Antenna Using Frequency-Independent Elements, IEEE Trans. AP, v41 n1, January 1993.
4. Frederick Emmons Terman, Radio Engineers Handbook, McGraw-Hill, 1943.

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

QUARTERLY DEVICES

New narrowband 10.7-MHz ceramic filters from Murata

Open the case of almost any entertainment or communication FM receiver, and you'll find at least one 10.7-MHz ceramic filter. Over the past two decades, these tiny piezo-electric devices have taken over as the IF bandpass filter of choice in consumer-grade radios. One reason for their popularity is price. In quantity, most ceramic filters cost less than a dollar. Another reason is performance; easy-to-install three-terminal filters deliver impressive selectivity, while lowering parts count, eliminating tuneup, and providing a convenient interstage matching impedance.

The one drawback normally associated with these filters is bandwidth; 10.7-MHz piezo-electric devices with passbands narrower than 100 kHz have been somewhat difficult to produce. This characteristic has prevented their use in many medium and narrow-bandwidth applications. However, Murata recently perfected an
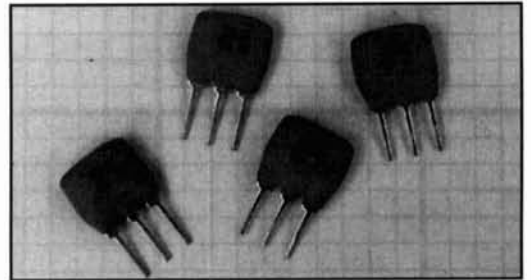
Specification of Ceramic Filter

Type: SFE10.7MT

1. Nominal Center Frequency (fn)	: 10.700 MHz
2. 3 dB Bandwidth	: fn ± 25 kHz min.
	Total 60 kHz min.
3. Selectivity at fn ± 100 kHz	: 20 dB min.
4. Insertion Loss	: 6.5 ± 2.5 dB
5. Spurious Response (within 9 MHz to 12 MHz)	: 30 dB min.
6. Ripple (within 3 dB Bandwidth)	: 1.0 dB max.
7. Temperature Coefficient of Center Frequency Temperature Range: –20°C to +80°C	: ± 20 ppm/°C max.
8. Input and Output Impedance	: 330 ohms
9. Withstanding Voltage	: 50 Vdc max. 1 minute max.

Figure 1. Spec sheet, mechanical drawing of SFE10.7MT.

Specification of Ceramic Filter

Type: SFE10.7MV5

1. Nominal Center Frequency (fn)	: 10.710 MHz
2. 3 dB Bandwidth	: fn ± 13 kHz min.
3. 20 dB Bandwidth	: 135 kHz max.
4. Insertion Loss	: 6.0 ± 2.0 dB
5. Spurious Response (within 9 MHz to 12 MHz)	: 35 dB min.
6. Ripple (within 3 dB Bandwidth)	: 1.0 dB max.
7. Temperature Coefficient of Center Frequency Temperature Range: −20°C to +80°C	: ± 20 ppm/°C max.
8. Input and Output Impedance	: 330 ohms
9. Withstanding Voltage	: 50 Vdc max. 1 minute max.

Figure 2. Spec sheet, mechanical drawing of SFE10.7MV5.

alternative process that allows production of narrow filters inexpensively.

Inside ceramic filters

To understand how narrowband ceramic filters were developed, let's look at how these devices work. The outer case, which is epoxy, provides a protective shell and tuned cavity for the vibrating ceramic element inside. The center operating frequency of this element is determined primarily by wafer thickness.

Passband characteristics are the product of several vibration modes that occur simultaneously within the filter's piezo-electric material. Design engineers can selectively control different modes by varying filter composition and by manipulating the shape of the electrode pattern deposited on the ceramic wafer. To optimize performance, vibration modes yielding the de-sired results are enhanced and modes that degrade performance are suppressed. The shaping of passband characteristics can become a complex process because altered variables tend to interact.

In the past, designers relied solely on *primary* vibration modes to produce usable 10.7-MHz filters. This approach yields the broad and relatively flat passband characteristics we normally associate with them (popular bandwidths range from 110 to 480 kHz to cover a variety of WBFM and other applications).

However, Murata engineers recently perfected a technique that uses *harmonic* vibration modes to produce filters with a much narrower passband response. This innovation has allowed them to expand the popular SFE10.7 product line into the medium and narrow-bandwidth filter category.

Presently, the newly developed harmonic-mode SFE10.7 filters come in two bandwidths. Complete specifications for each are shown in **Figures 1** and **2**. The SFE10.7MT provides a −3 dB bandwidth of approximately 50 kHz. The narrower SFE10.7MV5 provides a −3 dB passband of around 26 kHz. Both filters feature similar case size, pinout, I/O impedances, ripple, and insertion loss characteristics as other wideband SFE10.7 devices. Both may be used as drop-in replacements for other SFE10.7 filters where appropriate.

Narrowband ceramic filters (like all ceramic devices) are very easy to incorporate into IC-based receiver designs. This is because most semiconductor manufacturers now design the IF ports on their mixer and "receiver-on-a-chip" ICs for direct connection to standard piezo-electric devices. This practice virtually eliminates the need for adding external matching circuitry (see **Figure 3**).

Practical applications

The passband requirements for high-speed

Figure 3. Typical interface schematics with 2 or 3 common ICs.

Figure 4. Dual-conversion passband, wide versus narrow 1st-IF filter.

selectivity on the 455-kHz 2nd-IF filter; a practice that degrades the receiver's adjacent-channel selectivity in strong-signal areas (see **Figure 4**). Substituting a 50-kHz wide SFE10.7MT in the 1st-IF would decrease the severity of this problem—and installing a 26-kHz wide SFE10.7MV5 might virtually eliminate it!

2. Single-conversion receivers. There are a number of excellent single-conversion receiver chips on the market that, when paired with the right IF filters, offer new product opportunities. For example, the 50-kHz-wide SFE10.7MT should work very well for 136-MHz WEFAX, high-speed data links, telemetry, and WBFM wireless mics—to name only a few possibilities. The narrower 26-kHz SFE10.7MV5 should work well for single-conversion NBFM (voice and 9600 baud packet), airband AM, spread spectrum, and a host of other digital or analog applications.

For more information

You may obtain additional information on the SFE10.7MT and SFE10.7MV5 directly from the company. For full specifications, pricing, nearest distributor, and sample requests write: Murata North America, 2200 Lake Park Drive, Smyrna, Georgia 30080; or call: Casey Saunders at (404) 436-1300 or Harry Moore at (404)436-1300.

Designers should be aware that Murata manufactures a very wide range of communication-type piezo-electric filters, including 455-KHz narrowband ceramic filters and discriminators, several types of 10.7-MHz ceramic filters in lead and surface mount packaging, and a very attractive line of inexpensive SAW (surface acoustic wave) filters with popular VHF and UHF passbands. For information on the full line of Murata filter products, call 1-800-831-9172 and request the Piezo-electric Devices catalog number P-05-A.

Conclusion

The SFE10.7MT and SFE10.7MV5 are available now and they fill a rather large product gap in the current 10.7-MHz filter market. Hopefully this will open up new opportunities for some innovative receiver designs. Harmonic mode ceramic filters should work especially well in conjunction with many of the newer single-conversion receiver ICs. The SFE10.7MT and SFE10.7MV5 are available now, and I can't wait to try them out!

If you come up with a novel application or interesting construction project using this (or any) Quarterly Device, we'd like to hear from you. Please contact us by phone at (603) 664-2515, or send your manuscript to Communications Quarterly, P.O. Box 465, Barrington, NH 03825-0465.   ■

data, FAX, digital audio, and a host of other new communication modes do not necessarily fit within the boundaries of traditional message-channels we've used in the past. For this reason alone, the new MuRata devices should enjoy immediate popularity (they fill a huge product gap that presently exists between standard 12.5-kHz NBFM crystal filters and 100-kHz-plus WBFM ceramic filters). In addition to facilitating new modes, they should help us solve some lingering narrowband filtering problems in low-cost NBFM/FSK/AM/SSB designs (see below). Here are just a few ideas for immediate applications that come to mind:

1. Dual-conversion receivers. A huge number of scanners, handheld transceivers, and 2-way mobile radios employ dual-conversion IFs operating at 10.7 MHz and 455 kHz. To lower production costs, designers frequently specify a broad ceramic filter rather than a more-expensive narrow-band crystal filter to establish the 1st-IF passband. Unfortunately, this tradeoff places the total burden for message-channel

Howie Cahn, *WB2CPU*
295 Beacon Street #23
Boston, Massachusetts 02116
Internet: wb2cpu@world.std.com

THE FINAL TRANSMISSION

The Medium is the Message and Vice Versa

In the '60s Marshall McLuhan caused a stir with his statement "the medium is the message."[1] He meant that it is the *existence* of a communications medium, rather than the information it carries, that shapes how it's used and perceived. Early in the century, when movies were new, people would pack the halls regardless of what was being shown. Just seeing *moving pictures* was reason enough to go. The early days of amateur radio were similar. Then, using the new radio technology was an end unto itself. Being able to build and use the equipment to contact distant places was sufficient justification for operating. Sure, we'd occasionally do something useful like providing emergency communications, but in general the content of our transmissions was less important than the ability to make them. As communications technology matures, things are changing. Working DX is no longer very difficult and no longer impresses people. In fact, non-hams are usually amazed only by how unsubstantive our contacts are. With the novelty wearing off, we'll be less likely to use radio for its own sake, making random contacts just to see what locations we can reach. We'll need new things to do with our radios that take advantage of the exciting new technologies available, while still being interesting. The use and content of our medium will decide what amateur radio becomes, or, whether there's any reason for it to continue at all. McLuhan's dictum has been reversed—now "the message is the medium."

What we choose, or are allowed, to send will determine the technology we use, which transmission modes become part of our operations, and what types of equipment we develop and operate. Federal regulations restrict what and how we transmit. There are prohibitions,

among others, on commercial content, broadcasting, and on the modes and formats we can use. These rules are needlessly limiting the evolution of amateur radio. I would like to talk about why some of them are no longer required and should be modified.

To go from the abstract to the specific, let me give a personal illustration of what I mean. I use personal computers at work and at home. I carry a floppy disk back and forth with me to transfer files between the two machines. In spite of this, I often find the information I need is on the other computer or, that I'm not sure if the file I have on the machine in front of me represents the latest version of the data. I considered solving the problem by setting up a radio data link over the line-of-sight, five-mile path between the two sites. Then I realized how little of my data could be transmitted legally under the current amateur rules. The documents, programs, or spreadsheets that I produce for work clearly can't be sent. My checkbook files are probably OK, but, since it's illegal to encrypt them, transmission could involve sharing my personal finances with the world (similarly for copies of postal or E-mail correspondence). I couldn't even send this article if it was something for which I might get paid. Setting up the data link might be an interesting technical project, but it's relatively useless for the purpose I had in mind. I'm sure others have also been prevented from exploring ways of using our bands by these obstacles.

Of course, the content of our radio channels depends on what we're trying to do. Many traditional ham activities, such as message handling and DXing, are becoming either irrelevant or less interesting than they were. The service we've provided as a backup communica-

tions system for emergencies is being superseded by cellular phones and the coming satellite-based telephone networks. Ham contesting as we know it will ultimately be doomed by the diminishing pool of CW operators. The ham experimenters developing our digital networks, protocols, and TNCs are doing interesting work, but they are more concerned with building the equipment than the way in which it may be used. The appropriate question isn't only "What new technologies can we use?" but also "What can we do with radio technology that will get people enthusiastic and involved in using it?". If amateur radio is to survive as anything more than glorified CB or a refuge for nostalgia buffs, we're going to have to find some new purposes and goals.

I think our most important function is an educational one: providing a motivation to explore technology and a vehicle for doing so. This need will remain as long as having technically literate people is important. The obvious place to aim this effort is at young people. The flow of kids into ham radio has slowed to a trickle. There are probably several reasons for this, but apparently neither the activities nor the culture we present is of much interest to them. Let's face it, we're a pretty staid and conservative group—not exactly the most appealing combination to teenagers. We should look for new approaches to using radio, keeping what's still appropriate but being willing to incorporate changes as well.[2]

One potential role for us is as an adjunct to the coming "information superhighway" we hear a lot about. We could provide the "on and off ramps" either for low-cost access or for access at sites not wired into the network. Another possibility would be to change long term policy and allow non-commercial amateur broadcasting. To do either we'd have to change the rules. Already, packet networks are being limited in their ability to serve as Internet gateways by content restrictions. I think that the prohibition on broadcasting and limits on what you can send is a major reason why amateur television (ATV) is used by only a small number of hams. As the information networks move on to handling a full range of multimedia formats—video, music, animation, or whatever manifestations of cyberspace ultimately arise— we in the ham community will be left behind if we're not allowed to handle them.[3] The existing content and format restrictions will prevent amateur data networks from ever being more than a pale shadow of non-amateur networks like Internet and its descendants.

Many of the restrictions date back to the early decades of amateur radio. Then, those in the existing wire-based telephone and telegraphic services, as well as in the emerging

commercial broadcast industry, were legitimately worried about competition from unpaid amateurs. Format restrictions were partly the result of government fears about encrypted messages from foreign spies, first German and later Soviet. I doubt that executives of modern communications companies lose much sleep worrying about radio amateurs stealing business from them now, and the changed geopolitical situation, together with the sheer impossibility of tracking all the possible espionage channels, make the case for these regulations less compelling. Yet, not only are we still governed by these rules from sixty or seventy years ago, they've become so much a part of the amateur tradition that we accept or even welcome them. They've been internalized to the extent that you may even feel angry and threatened when I suggest we be given more freedom. Some of the restrictions are still valid, but we should start from scratch and see which ones are relevant and why.

The major need for any content and format restrictions now is to protect amateur instead of commercial users. I certainly wouldn't want a pizzeria or taxi company deciding to use two meters for dispatching their drivers (with an overly cooperative VE on the payroll "helping" employees get licenses). Similarly, I wouldn't want to see someone transmitting wideband ATV on twenty meters. Nothing I'm recommending would alter the essential, non-commercial nature of amateur radio, but there are changes that could be made that would broaden its attraction.

The prohibition on broadcasting seems particularly unnecessary and limiting. Being able to broadcast is something that seems to excite and empower people, particularly kids. Recent movies like *Pump Up the Volume*, about a teenage pirate radio broadcaster, or even *Wayne's World* demonstrate this appeal. Let me make a controversial suggestion: that broadcasting and the transmission of music be permitted on one UHF band, say, 13 cm (2.3 GHz). This may sound like an absurd and scary idea evoking images of teenagers emulating their favorite MTV VJs and playing *Mega-Death* videos, or worse. While it's true that some would use the opportunity to do things of questionable social value, at least it might get kids psyched up enough to learn about and use radio (and related technologies) and pull them within the framework of amateur radio. Some might go on to develop digital networks. Some might become HF operators. Some would go on to careers in technology-based industries, like many of my teen ham friends of thirty years ago. Broadcasting implies having listeners or viewers. They'd also be introduced to ham radio, in a form to which they would

relate. Ham broadcasting could also provide a real service, true public access to the air waves. We have more commercial broadcast and cable channels now, but most offer variations of the same programming. Even the re-quired public access cable channels are available only to particular groups (which groups depend on the biases of the specific community). Ham broadcasting would, for the first time, give access to all. Maybe none of these positive things would happen; but, what's the harm in letting interested parties transmit with a watt or so and an omnidirectional antenna, on an un-derutilized band we're in danger of losing anyway?

I've described a series of paradoxes and dilemmas that characterize where I think we amateur radio devotees stand. Communications technology has matured to the point that merely making contacts is no longer very interesting. Advanced new techniques and systems are coming along, but we haven't discovered any-

thing to do with these systems that will appeal to many beyond those directly involved in creating them. We've failed to adapt our mission to the new circumstances and, as a result, we're not attracting bright kids as we have in the past. I think that the current regulations needlessly inhibit exploring new roles. Maybe I'm wrong. Maybe radio has progressed to being a tool or appliance, like a dishwasher or telephone, and no longer warrants having a hobby based on using it. Maybe hams are inherently so shallow and unimaginative that, even given all the freedom in the world, we still wouldn't do anything interesting. For now, I won't be *that* cynical. What we need are some creative ideas on where to go from here in using our medium. ∎

REFERENCES
1. Marshall H. McLuhan, *Understanding Media*, McGraw-Hill, 1964.
2. Howie Cahn, "Bringing Amateur Radio into the Computer Age," *Communications Quarterly*, Winter 93, page 94.
3. Bryan P. Bergeron, "Multimedia Communications—the Future of Amateur Radio?," *Communications Quarterly*, Winter 94, page 13.

Corrections

Who in the World is Peter X?

At the end of the sidebar "The Planet Mars QSO Cup" that appears in the article "HF Radio on Mars" by Craig Fry and Robert Yowell (Spring 1994), you'll notice that the author's name is given as Peter X. Of course, the author is really no other than our Senior Technical Editor Peter Bertini, K1ZJH, and he informs me that he hasn't changed his name recently to become a pirate radio station operator, or anything of the sort! Sorry Peter, I'll try

to keep my fingers in the right "pew" of the keyboard next time I type in your name! Ed.

Problems With Parts Placement

In "A Gyrator Tuned VLF Receiver," by Arthur Stokes, N8BN, (Spring 1994), there were some errors in parts placement in **Figure 4** on page 26. Capacitors C7 and C8 should be connected to the groundplane of the board, and diodes D3 and D4 should be interchanged. A corrected drawing is included here. Ed.
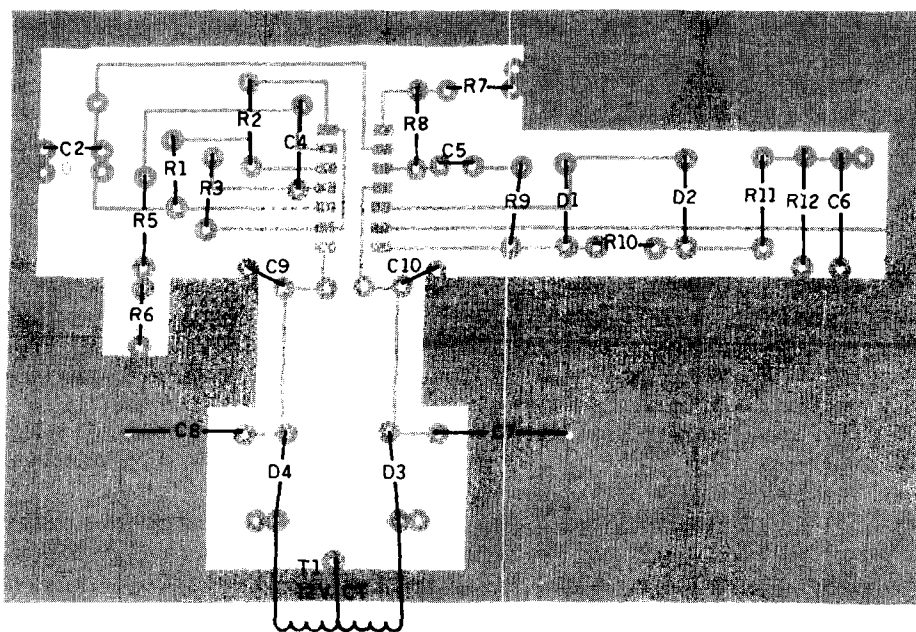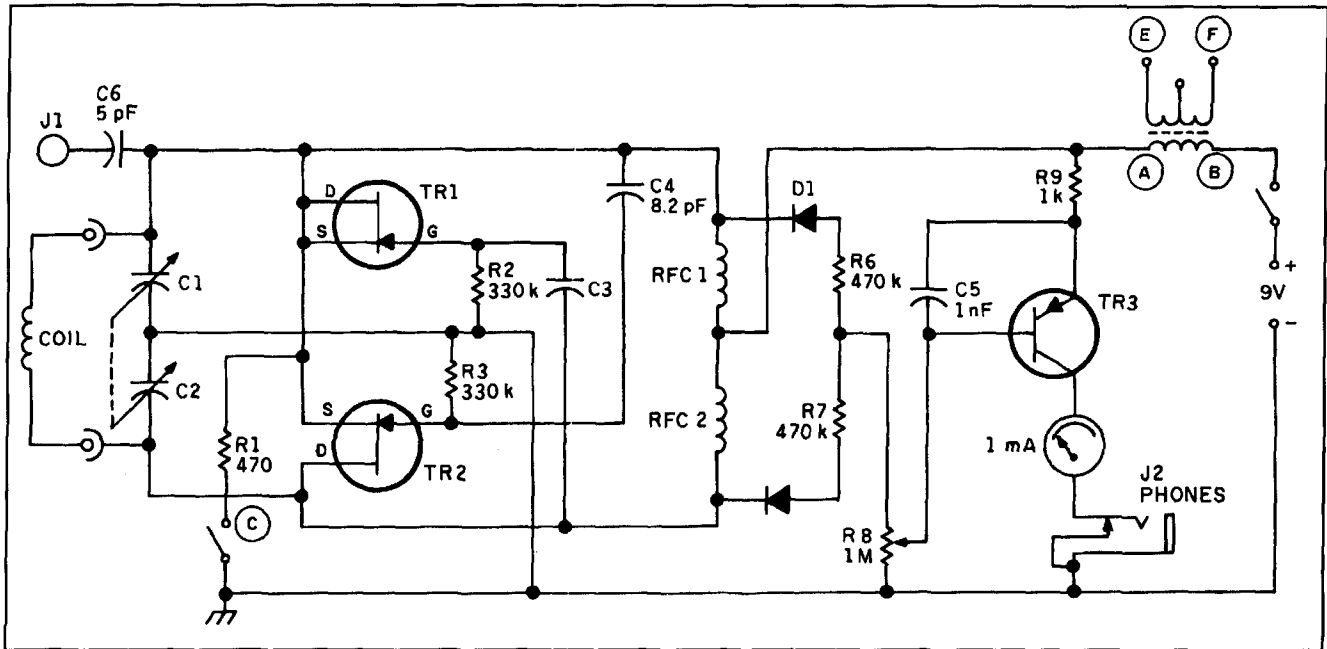
Figure 1. Circuit diagram of the dual FET dip oscillator.

PRODUCT INFORMATION

Tektronix Expands DMM Line

Tektronix Inc. introduces three new hand-held digital multimeter (DMM) products for use in applications ranging from industrial design and service to education and the general hobbyist.

The three instruments are the DM255, DM256, and DM257 general-purpose multimeters, priced at $65, $75, and $85 respectively. Each is designed and manufactured to comply with safety standards established by UL and IEC and conform to Mil-T-28800, Class 2 standard for shock and vibration.

The DM255 features a 3-1/2 digit display, 0.7 *percent accuracy, autoranging or manual selec-*tion, data hold and current measurements. Where current measurements are not needed, the DM256 offers testing of voltage, capacitance, resistance and diodes with 0.5 percent accuracy, a fast continuity beeper and memory offset. The DM257 offers current and capacitance measurements with 0.5 percent accuracy, autoranging and manual selection, and a low battery indicator.

For more information, contact Tektronix Electronic Service Tools, Tektronix, Inc., P.O. Box 1520, Pittsfield, MA 01202 or call Tek Direct, the Tektronix national order desk, at 1-800-426-2200, extension 183.**Rail-To-Rail**

AEA Satellite Tracker

Advanced Electronic Applications, Inc. introduces their ST-1 Satellite Tracker. The ST-1 offers hands-off control of antennas and tranceiver tuning to make satellite operation easier. Features include a completely resident program (TSR) that runs concurrently with other programs without interfering with foreground programs; automatic control of Yaesu 5400/5600 Azimuth-Elevation rotors. The ST-1 also has A/D converters, precision power supply, and relays to control two independent rotors. Works with InstantTrack, QuickTrack, and RealTrack software; tunes FT-736, TS-790, IC-970, and the IC-475/2785 combination transceivers. Hardware uses one serial port and parallel port from the host computer. Auto-matic switching provides easy access to ports when the ST-1 is not in use. Works together with other satellite software, like PG_AEA, PB, and PG. Allows your satellite packet station to be fully automated and run the PACSATS unattended.

Suggested retail price for the ST-1 is $399. The ST-1 is available from your favorite amateur radio dealer.

For more information, contact Advanced Electronic Applications, Inc., P.O. Box C2160, Lynnwood, WA 98036, Phone: 206-774-5554; Fax: 206-775-2340.

Coil	Number of turns	SWG	Length	Coverage (MHz)
A	3	22	4 mm	23–60
B	10	22	12 mm	13–29
C	20	22	10 mm	6.6–14.5
D	35	26	12 mm	3.6–7.5
E	100	26	32 mm	2.0–4.1
F	DC antenna coil on 9x40 mm ferrite rod; remove turns to adjust.			1.1–2.2

All coils wound on 16 mm OD formers.

Table 1. Coil data.

grounding point "H" (see **Figure 3**). This feature is very useful when checking receivers.

Construction Notes

The circuit shown in **Figure 1** covers from 10 to 60 MHz and there is very little change in meter reading as C1C2 is tuned across the band. The size of C1+C2 determines the coverage of each coil. I originally used a 360+360 pF gang from an old receiver which was fine for the lower frequencies, but the scale was too crowded on the higher frequencies. The 220+220 pF gang (also from an old radio) offered a good compromise; it also had an integral reduction drive, and I consider a slow motion dial a *must* for dip oscillators.

Coil winding details are given in Table 1, but they are only a guide. Start with the lowest frequency band and adjust the winding of each coil so as to provide some overlap at the higher end of each range. I used some two-pin formers from an old Eddystone wavemeter, but coil formers can be made from narrow PVC conduit, discarded plastic syringe barrels, or even purchased!

Obtaining an observable dip on the lowest (1 to 3 MHz) range proved to be difficult until I hit upon the idea of using a short (30 mm) piece of ferrite antenna rod in the coil. The original broadcast antenna coil was carefully "pruned" until I obtained the right coverage. This low frequency coverage, not usually found on most dip oscillators, has proved very useful for my interests in restoring old communications receivers.

I constructed my dip oscillator on two scraps of Veroboard (see **Figures 3** and **4** for layout). Alternatively you could use tag boards, but do ensure that the RF oscillator section has a symmetrical layout. Make sure that the leads between the coil socket, tuning capacitor, and oscillator section are short and rigid—I used 1 mm wire. The meter amplifier was build on a miniature tag-strip mounted on the positive meter terminal; the meter may be anything from 100 μA to 1 mA—surplus VU meters are very suitable.
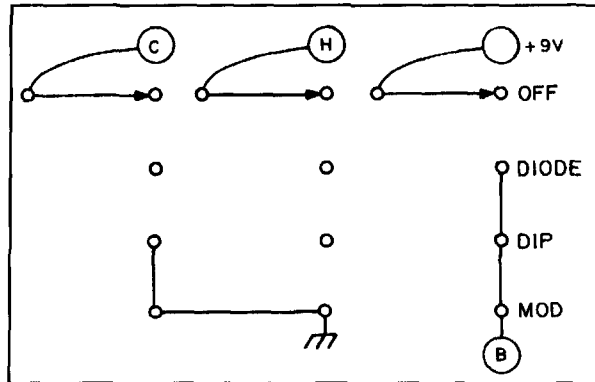
Figure 2. Function switch (3-pole, 4-way).
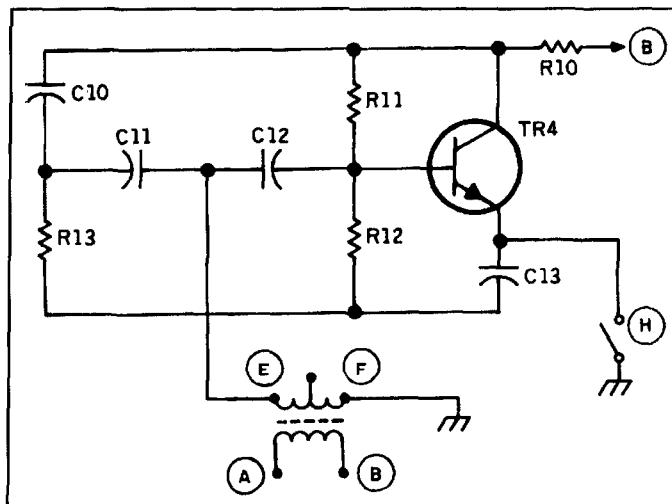
Figure 3. AF oscillator/modulator schematic.
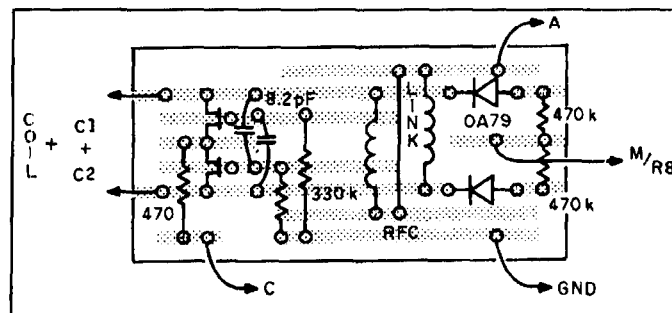
Figure 4. DIP oscillator board..
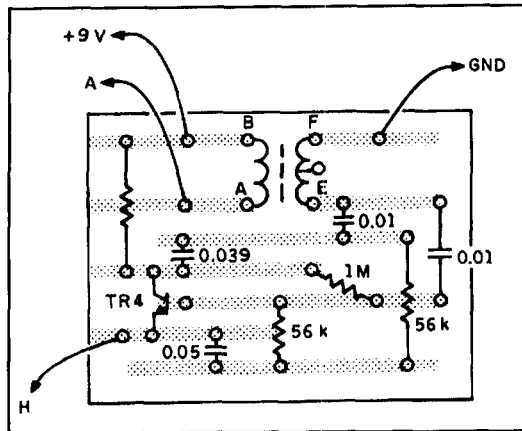
Figure 5. AF oscillator/modulator board.

If the unit fails to oscillate over the entire range, try increasing coupling capacitors C3 and C4 and/or altering the values of the two RF chokes to somewhere between 100 µH and 1 mH.

The modulation transformer is a miniature input circuit type and should have reasonably high DC resistance across both windings (not

Parts List

Oscillator Section

C1,C2	220+220 pF (tuning, see text)
C3,C4	8.2 pF
C5	1000 pF
C6	5 pF
R1	470
R2,R3	330 k
R6,R7	470 k
R8	1 M linear (sensitivity)
R9	1 k
TR1,TR2	2N5486, MPF102 or similar
TR3	BC 177 (pnp) or similar
D1,D2	OA91 or OA85
RFC1,RFC2	100 µH

AF Oscillator/Modulator

C10	0.039 µF
C11,C12	0.01 µF
C13	0.05 µF
R11	1 M
R12,R13	56 k
R10	39 k-56 k (adjust for modulation level)
TR4	BC108 or similar

All transistors are 1/8 watt.

less than 50 ohms). Jack J1 serves to couple a short pick-up antenna for use as a field strength meter in the DIODE mode. It is also used to couple to a digital frequency meter for accurate frequency measurements.

REFERENCES
1. H.L. Gibson, *Test Equipment for the Radio Amateur*, 2nd edition, RSGB, 1978, page 3.13.
2. G.R. Jessop, *VHF/UHF Manual*, 4th edition, RSGB, 1983, page 10.21.
3. *Electronics Today International*, June 1975, page 38.
Captions:

In our final note for this issue, Marv Gonsior, W6FR, offers a modification to the Titan 425 power amplifier that provides additional over-drive protection.

Titan Overdrive Protection

Marv Gonsior, W6FR

Ten-Tec, manufacturer of the popular Titan 425 power amplifier, recognized the necessity for overdrive protection to preserve one's investment in the expensive 3CX800A tubes that have a grid dissipation of only 4 watts. Prudent operators will surely do the same, since obviously, modern exciters are easily capable of decimating those fragile grid structures. Ten-Tec has incorporated a red warning LED, ALC, and a foldback circuit internally, but these features may be overcome by sheer power along with the failure to observe the warning light and/or the grid current meter. Here is a simple way to augment the existing features that will more effectively warn an operator of overdrive. The modification is extremely modest in cost, and doesn't involve any cutting or disruption of the existing circuitry.

The manufacturer-installed red warning LED has a measured 1.6 pulsating DC volts on its anode when conducting. This provides a suitable voltage source to actuate any number of warning devices and lock-out relays.[1] However, I felt that the combination of a "screamer" plus an optional flashing red LED would be a relatively cost-effective means of warning to preclude overdrive—especially when it's blasting away and blinking in my face, or when the amplifier is situated out of my field of vision.

The circuitry, shown in **Figure 1**, requires only two alkaline 9-volt cells for power. These batteries will easily last their shelf life, assuming you're a fast learner! All components are off-the-shelf at Radio Shack. Before beginning any work, a word of warning. *High voltage*

kills! Be sure the high voltage cable is disconnected from the amplifier, and that the high-voltage capacitors have had sufficient time to discharge. Take this opportunity, while you have the lid off the cabinet, to clean out any dust or debris that may have accumulated inside the amplifier.

Now, connect a wire that's soldered to the anode of D1 on the Display Board # 81220, and shown in the owner's manual in **Figure 5-25**. The anode connection is easy to recognize, as the cathode is grounded to the foil. I used a small Teflon™ insulated wire that I slid easily through the wiring harness to the rear of the chassis and out through one of the unused phono plugs on the rear panel. If you don't wish to use the 9-volt cells, you can steal the plus 28 volts DC from the power cable at the rear of the cabinet. Use it accordingly with a suitable dropping resistor, as I first did. However, I decided that an internal battery power supply was preferable to avoid any potential complications. The common lead is attached to the ground lug at the rear of the amplifier cabinet and the two leads are terminated in a RCA type phono plug for connection to the rear of the unit.

In my particular application, I used an extremely small double-pole, double-throw hermetically sealed surplus relay from my junkbox. It is actuated by a control voltage of approximately plus 0.8 volts DC; (see **Figure 2** for the schematic.) The relay obviously contains just an ordinary NPN switching transistor. The schematic shown in **Figure 1** performs the same function and is probably much easier to duplicate.

The Radio Shack "magnetic buzzer" doesn't sound like one at all because its frequency is given as 3.4 kHz and is specified to operate from 6 to 16 volts DC, drawing about 10 mA. Of course, its output level is dependent upon its supply voltage. The price was very attractive, the performance was adequate, and the part was readily available. For very intermittent service such as this, a plus V of 18 volts won't harm the unit and, of course, provides a desirable few extra dBs of output!

I mounted my "squawk box" in a small aluminum enclosure, but any type will do because there's nothing critical about any of the components. Radio Shack's # 270232 might be a suitable candidate if you start from scratch. Mine came from the junkbox.

To test the unit, give one quick whistle with a bare amount of overdrive, as verified by diode D1 in the amplifier, and stand by for an instanta-
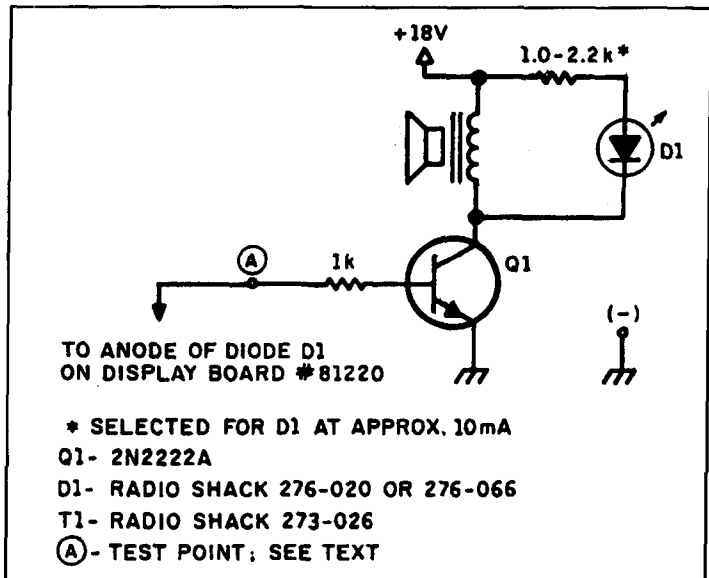
Figure 1. Overdrive warning circuit.
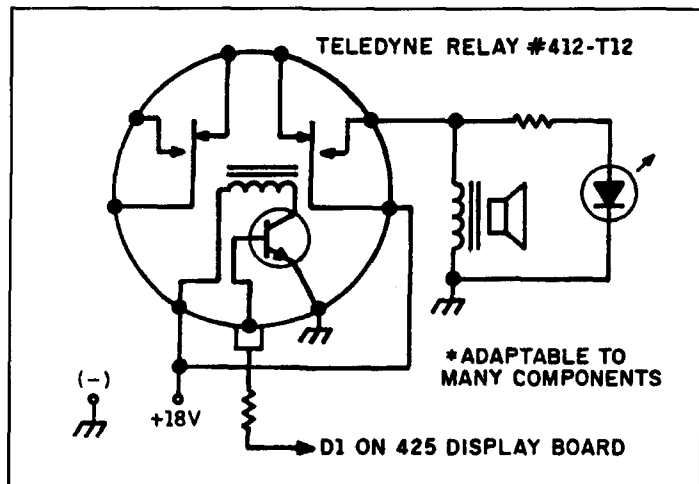
Figure 2. Alternative overdrive warning circuit.

neous response! Alternatively, place a plus 1.5 voltage source to point "A" as a bench test.

This device is an extremely inexpensive insurance policy with a very small, one-time premium. It's one that you don't ever want to hear or see, but awful comforting to know you have around! A final note, if you use batteries, don't forget to check and/or replace them once a year, or so, when you do your smoke alarms. I placed two small pin jacks on the box, so the plus V can be checked with ease periodically.

REFERENCE
1. Henry Gow, "Grid-Current Trip Circuit for the Titan Linear Amplifier," *QST*, June 1993, page 74.

LETTERS

Loop transmission is older than originally described

Transmission using loop antennas is even older than that described in the 1928 clipping sent in by L.B. Cebik, W4RNL.

The oldest reference that I can find is from the February 1920 issue of *QST*. This describes a portable transmitter, spark of course, used by the army during WW-I. It used a three turn loop one meter square. It was manufactured by the Wireless Specialty Apparatus Co.

There are several articles about loop transmission for hams scattered through various issues of *QST* through 1923.

Harry R. Hyder, W7IV
Tempe, Arizona

Thanks! ... and by the way ...

I wish to thank you for publishing my article on the QRP rig. I hope it will be of interest to many of your readers. I did notice one thing that probably should be noted.

The panel layouts on page 34 were drawn to a 1:1 scale but printed at approximately a 2 to 1 reduction (1.94:1 to be exact). No indication was given as to the scaling but one could determine this from the chassis drawing. One has to copy this on a copy machine onto the transparency. I am sure builders can figure this out on how to increase its size.

Thanks again.

Jim Pepper, W6QIF
Orinda, California

Anyone wishing to build W6QIF's deluxe QRP station can obtain printed circuit boards from FAR Circuits. 18N640 Field Court, Dundee, Illinois 60118. Prices are as follows: receiver board, $5.60; transmitter board, $4.40; counter board, $3.50; side-tone board, $3.50. A set of boards is $14. Please add $2 shipping and handling per order.— Ed.

Let's get real

I just received the winter 1994 issue. Eeee. Leaning forward! Bryan Bergeron's articles have always had a "progressive air" about them, but combined with other things I'm seeing in CommQ, I begin to wonder.

I read with trepidation your editorial in the same issue. My GLR (Gut Level Response)? Everything is turning into virtual this or virtual that! Notice even the photograph in Bryan's article on virtual tools is virtual. In fact, the last year of CommQ has shown Bryan's surrealistic unrealities on each issue's front cover. Have you seen *Midnight Engineering's* recent cov-

ers? Same stuff, this time with a theme of fractals. You mention virtual reality mobile operation. But I thought the anxiety, adrenaline, sideshows, camaraderie, and successes were what made an expedition, or a Field Day! Where has *real* reality gone?!

Communications Quarterly seems to be increasingly leaning toward a sci-fi future. I guess somebody needs to do this to get the practical people thinking. Is this drift in style and content intentional? I'm not against movement in the computerized direction. I met Howie Cahn, WB2CPU, over the Internet prior to his publication in CommQ. I looked forward to his article in the Fall issue ("Connecting Computers to Radios," page 13) and was not disappointed. His sense of balance is keen. The article was well done and I told him so.

Pages 52–55 of the Winter '94 issue also surprised me. Robert Witmer used four pages to say what could have been communicated with a handful of formulas. Didn't a reviewer question this? For example, consider the first half of the table on page 52/53:

increment = (freq-222) * 100
set = int (increment/10)
EPROM address = increment/2 = 3 * set

The sidebar on page 56 does a good job of covering the hex<>decimal conversion process. So why did we need two columns of decimal and hex numbers in all four tables?

See also Michael Gruchalla's article in the Fall 1993 issue. Twelve pages of plots were not necessary. If he summarized what he found with maybe two or three plots as examples, I would have believed him, really. I trust an author to look at the gory details and present to the reader a consolidated summary of the work s/he did.

I am not advocating that the bare minimum is best for understanding (ever read an APL program)? I've tried to keep a balance. Do you still have copies of *Ham Radio* around? Look at pages 34 and 38 of the November 1987 issue. I tried to show construction details in **Table 1**, and convey understanding with **Table 3**. Maybe I, too, went overboard with **Table 1** and could have just described it in text. I did better on pages 60 and 61 of the April '88 issue. The formula conveys all that's going on, while **Table 2** gives the quick reference for those who want to just get on with operation.

I think it's important for the authors themselves to get feedback on their work. I know I appreciate it. I normally send copies to all relevant parties when I write a letter like this, but

because, in this case, it's more of an overall impression of where CommQ is heading, I sent it only to you. Pass on any information you feel is appropriate.

Brian Mork, KA9SNF
Fairchild, Washington

I liked it, but please make this correction

I found the short article on "Yagi Gain versus Boom Length" by David Barton, AF6S (Tech Notes, Winter 1994) very interesting! However, the boom length for the HG215DX should be 4.03, not 8.40. This will also make a difference in his graph. We agree totally on his conclusions.

Also, Brian Beezley uses a gain figure of merit based upon one of three formulas to determine the maximum possible gain. I like the K1GQ formula for HF beams:

$$G = 3Ln(B) + 9.85$$

This formula produces excellent results for most HF antennas. Exceptions are those with radiation resistance less than 10 ohms and those with high F/B (>30 dB).

Roger Cox, WBØDGF
Telex Communications, Inc.
Lincoln, Nebraska

Thanks for the assistance

I am afraid that this is a very late letter of thanks and acknowledgement for the additional information you provided in response to my request of earlier this year to Peter Bertini.

I had hoped to construct the meter by this time so that I might have included some comments about operating it in my letter to you. Sadly, however, the parts are still in their box, other pressures having intervened.

My thanks once more to you and your colleagues. Keep up the good work. The *Quarterly* is first rate.

John Teague, G3GTJ
Somerset, England

EDITORIAL *(from page 4)*

tors to install screening software into their systems that will flag and pigeonhole messages from certain amateur callsigns. The sysop can read these messages later, pass them on if they aren't in violation of the rules, or drop them from the system if they are. Unfortunately, because the operator tells the software which callsigns to flag, this system will only hold up transmissions from known offenders, so it's not foolproof. Messages in violation of the rules may still get into the network, despite the sysop's precautions.

What's a ham to do? Amateur radio has a history of being a self-policing fraternity. If the operator of the first forwarding station delays the transmission of information through the automatic packet system until he has time to review all the messages, he will be able to intercept any inappropriate messages. On the other hand, those down the line will have to wait for their mail until the sysop has a chance to screen it, and technology takes a back seat.

How, I wondered, do devoted packeteers feel about the latest steps taken by the FCC? To find out, I polled a few packet operators for their opinions.

The majority of people I spoke with felt that the FCC still hadn't gotten it quite right with their latest ruling on automatic packet. The consensus was that even though the originating station would be held responsible for his actions, an essentially innocent party—the operator of the first forwarding station—might still have to pay for the transgressions of another. At the same time, however, the control operators of voice repeater systems were absolved of their responsibilities in similar situations. A double standard?

The feeling still lingered that the FCC rules concerning message-forwarding systems continue to put a hold on amateur packet radio research and development. The need to put messages on hold, said one operator, takes away one of the most important niceties of the automatic packet network—the immediate transfer of information throughout the system.

There was, however, one operator who was unfazed by all the hoopla. She contends that it doesn't take too long to review the messages left on her board, and all messages are held until they are read. Furthermore, users of the board are identified by callsign and zip code when they log on; so if someone sends a message in violation of FCC rules, he can be reported to the authorities and tracked down with ease. My lone, unworried operator feels

that most board users don't want to hurt their sysop, so chances are that most users will obey the rules and avoid business communications and obscenities. "I think it's important to keep the networks clean in order to encourage young people's participation in packet. Controversial information is fine. I don't want to be a censor; however, I think offensive material and vulgar language doesn't belong. The way I interpret the rules, if I receive a message that violates FCC regulations, I can help the commission track the originator down." For this amateur, the new rules become a way to help maintain the integrity of the airways and not a hindrance to operation.

The new rules on automatically forwarded packet transmissions are a step forward, it's true. No longer will many pay for the sins of one. However, one person (the sysop of the first forwarding station) still must share the blame for any illigal transmissions he lets pass. The number of messages in violation of FCC rules should decrease—if those in charge of first forwarding stations do their part. However, given the hectic lives most of us lead, it's a safe bet that many BBS operators may choose to just let the mail flow down the information superhighway, unread.

Down the years, as amateur radio has changed and incorporated new technologies, there has been an ongoing need to redefine the rules governing the way hams operate. The FCC has taken a first step towards placing the blame for automatically transmitted packet messages that contain rules violations where it belongs; but they haven't gone far enough. Yes, with the exception of a stop at the first forwarding station, information is free to flow through the network. However, it would be a real step forward if the commission were to recognize that amateurs—as individuals, clubs, or informal collectives—should be allowed to serve their hobby as "common carriers" of information—much like the U.S. Postal Service or the telephone company.

When mail fraud is committed, we don't prosecute the postman for delivering the mail. When someone makes an obscene phone call, we don't prosecute the operator for putting the call through. *We blame the perpetrator!* Why then should the automatically controlled packet network be any different? It's time to ask the FCC to put the blame where it belongs—squarely on the shoulders of the person whose message is in violation. The new rules are a beginning, but we need more!

Terry Littlefield, KA1STC
Editor

PRODUCT INFORMATION

New Soldering Iron/Torch Kit From Philips ECG

Philips ECG announces that its Butane-Powered Soldering Iron/Torch is now available in a 10-piece kit (J-500KT) complete with carrying case, 4 different tips and other accessories.

The ECG® J-500 Soldering Iron/Torch is a hand-held tool that features a see-through refillable butane tank, a temperature control for accurate adjustment of tip temperature, and a built-in igniter.

The kit includes the J-500 soldering iron/torch, a solder tip, blow torch tip, hot knife tip, heat blower tip, metal safety stand, a cleaning pad, spare flints and a half-ounce of 60/40 rosin core solder. Everything is encased in a heavy-duty, fitted carrying case. Additional tips are also available. Suggested list price for the J-500KT is $35.50 and is available from participating Philips ECG distributors in the United States and Canada.

For more information, contact your local Philips ECG distributor.

Custom Software Rights In New Software Legal Book

Paul S. Hoffman, the author of The Software Legal Book, is a software lawyer and CDP with more than twenty five years of experience with software related contracts.

The book covers all major aspects of software ownership and contracting including such topics as ownership of copyright in software modifications; what software copyright protects; tools for analyzing 'rights' situations and determining who has the right to license; and form language examples for allocating custom software rights between parties. Also covered are: employee agreements; "work for hire;" "joint works;" "derivitive works;" international software protection; bundled system contracts; disclosure agreements; software distribution agreements; and handling software contracts in buying or selling businesses. A comprehensive separately tabbed index provides easy access to text discussion and sample clauses.

The appendix includes software-relevant portions of the copyright law and regulations and the Uniform Trade Secret Act, international copyright treaty tables, the Uniform Commercial Code, the 1988 bankruptcy law changes and the 1991 European Software Directive. Forms provided range from employee agreements to agreements for disclosure, custom software development, beta test, software distribution, teaming, and source code escrow. All form examples include some explanation or context information. As an option, selected forms are available on 3 1/2 inch disk in WordPerfect® format.

The 1994 book may be ordered from Shafer Books, Inc., P.O. Box 40, Croton-on-Hudson, NY 10520-0040 (phone 914-271-6919 or FAX 914-271-5193) at a cost of $185 (US); $225 with Forms on Disk; plus shipping. (10% discount on prepaid orders.)

Voltronics RS Series Ceramic Trimmer Capacitors

Voltronics Corporation RS series ceramic trimmer capacitors are now stocked in depth to meet increasing delivery demands.

The RS is 0.2 inch in diameter with two axial leads on 0.15 inch centers. It has three capacitance ranges 0.6 to 3.0pF, 1.0 to 5pF and 2 to 10pF. Its DC voltage rating is 50 with 100 volts DC withstanding voltage. TC is -500 ±1000ppm/°C. Q is over 250 at 1 MHz.

Prices are 46¢ each at 10,000 quantities with delivery from stock for quantities up to 5,000.

For more information, contact Nicholas J. Perrella, National Sales Manager, Voltronics International Corporation, 100-10 Ford Road, Denville, NJ 07834. Phone: 201-586-8585; FAX: 201-586-3404.

On-Board Guidance System

~Dual

~Dual

Scan Mode
Time
17

New
TH-79A
FM DUAL BANDER

Information at your fingertips. Everything you need to know about operating the new TH-79A FM dual-bander (144MHz/440MHz) can be viewed in its unique dot-matrix LCD with alphanumeric display. No need for the manual. In addition to this innovative guide function, the TH-79A sports a user-friendly menu system, providing easy access to the many powerful features of this slim-line handheld transceiver. Such as 82 non-volatile memory channels with ID, DTSS and page functions, and a DTMF memory function for auto-dial operation. Full-crossband duplex operation is available, as is the ability to receive two frequencies on the same band (VHF+VHF or UHF+UHF) simultaneously. And thanks to the FET power module, long hours of operation are possible on one charge. With the TH-79A, transceiver technology enters the 21st century.

Features

■ 2.7W approx. output (144MHz), 2W approx: output (440MHz) from MOS FET power module and supplied 6V battery; 5W approx. output using optional PB-34
■ Dot-matrix LCD with menu/guide system
■ 82 non-volatile memory channels with ID
■ DTMF keypad with memory function
■ DTSS (Dual-Tone Squelch System) with page
■ Built-in CTCSS tone encoder/decoder
■ Automatic band change ■ Power-on call sign display
■ Auto repeater offset (VHF) ■ Input overvoltage warning
■ 3-position output power control
■ Auto power-off and battery save function ■ Time-out timer
■ Multiple scan modes ■ Cross-band repeater function
■ Page answer-back function ■ Channel display function
■ Wideband receiver coverage, including AM receive on the aircraft band*
■ Modifiable for MARS/CAP use**

*Specifications guaranteed for Amateur bands only.
**Permits required. Specifications guaranteed for Amateur bands only.

KENWOOD COMMUNICATIONS CORPORATION
AMATEUR RADIO PRODUCTS GROUP
P.O. BOX 22745, 2201 East Dominguez St., Long Beach, CA 90801-5745
•Customer Support/Brochures (310) 639-5300
•Repair Locations/Parts (800) KENWOOD •Bulletin Board Service (BBS) (310) 761-8284
KENWOOD ELECTRONICS CANADA INC.
6070 Kestrel Road, Mississauga, Ontario, Canada L5T 1S8

KENWOOD

94ARD-0970