$5

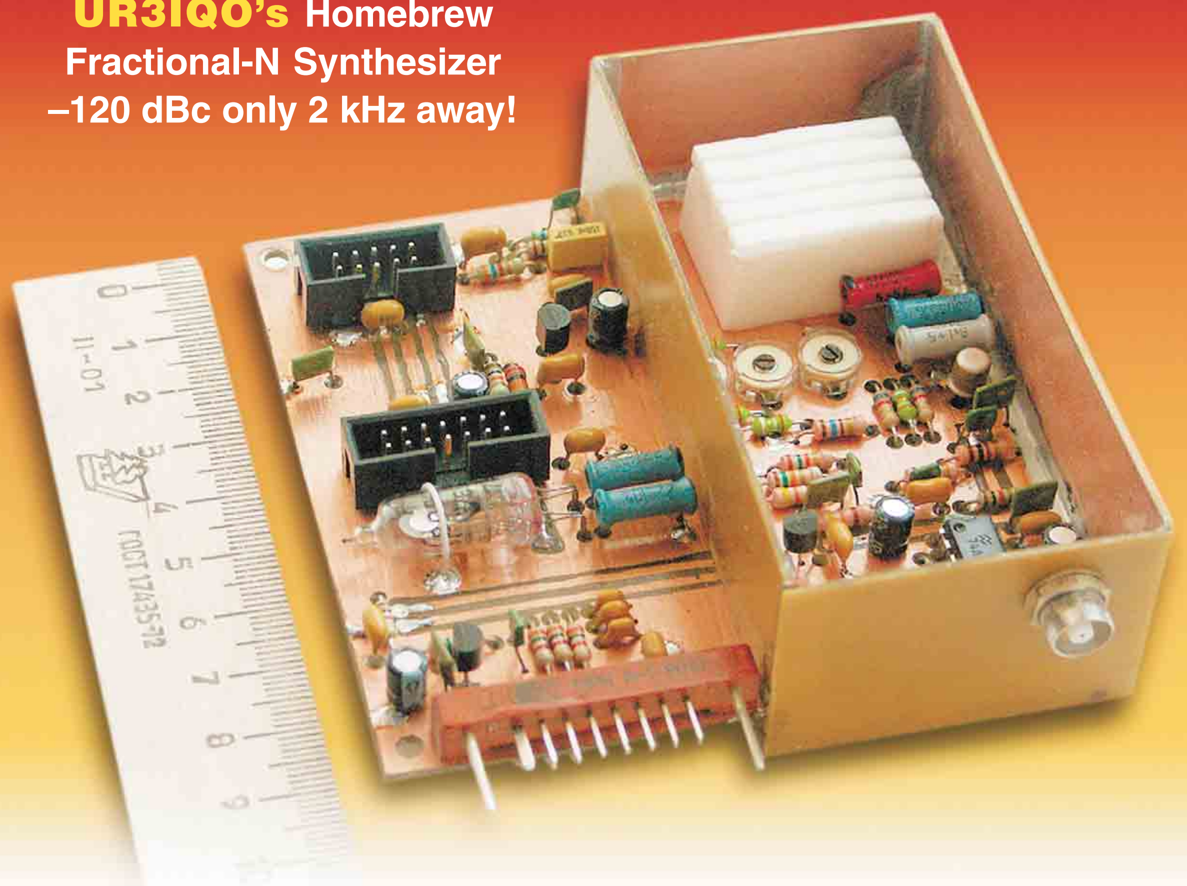
QEX

INCLUDING: COMMUNICATIONS QUARTERLY

Forum for Communications Experimenters

November/December 2003
Issue No. 221

UR3IQO's Homebrew Fractional-N Synthesizer –120 dBc only 2 kHz away!



ARRL The national association for AMATEUR RADIO
225 Main Street
Newington, CT USA 06111-1494

INCLUDING: COMMUNICATIONS QUARTERLY

QEX (ISSN: 0886-8093) is published bimonthly in January, March, May, July, September, and November by the American Radio Relay League, 225 Main Street, Newington CT 06111-1494. Periodicals postage paid at Hartford, CT and at additional mailing offices.

POSTMASTER: Send address changes to: QEX, 225 Main St, Newington, CT 06111-1494
Issue No 221

Offices

225 Main St, Newington, CT 06111-1494 USA
Telephone: 860-594-0200
Telex: 650215-5052 MCI
Fax: 860-594-0259 (24 hour direct line)
e-mail: **qex@arrl.org**

Subscription rate for 6 issues:

In the US: ARRL Member $24, nonmember $36;

US by First Class Mail: ARRL member $37, nonmember $49;

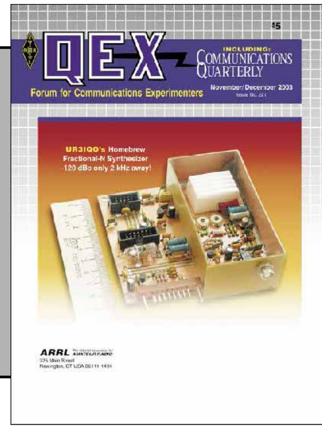
Elsewhere by Surface Mail (4-8 week delivery): ARRL member $31, nonmember $43;

Canada by Airmail: ARRL member $40, nonmember $52;

Elsewhere by Airmail: ARRL member $59, nonmember $71.

Members are asked to include their membership control number or a label from their *QST* when applying.

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



About the Cover

High-performance from UR3IQO's workbench. The story begins on p 25.

Features

Columns

Nov/Dec 2003 *QEX* Advertising Index

Empirical Outlook

Old Technology Come New

Recently, I had my television receiver on and I came across Larry King. One of his guests was remarking about how Amateur Radio had played a significant role in the aftermath of the September 11, 2001 disaster. "It's old technology, but it still works," he said.

Well, yes, radio is old technology by now—but I'm not sure that's what the commentator meant. I think what he meant was that Amateur Radio technology is old and that nothing is new under the sun.

My friends, we must change that perception. We have a chance—and we must take it—to demonstrate for our government officials and for the world just what we are accomplishing. We call for volunteers to write and demonstrate to the FCC and others just how we are fulfilling our purpose of technical innovation.

As in so many endeavors, ours boils down to a marketing issue. Our Service is supposed to be something that serves the public. However, that is not how it is generally perceived. It is perceived as a hobby with a bunch of guys and gals playing around with radios for pure enjoyment. Over the years, I've learned that you have to acknowledge general perceptions, then proceed to destroy them when they are wrong.

How can we do it? Write a letter. File your comments with FCC under their ECFS (electronic comment filing system) on the Web. Contact your congressman or senator via e-mail. Get involved.

One other basic purpose of our Service is to promote international goodwill. We are pleased that *QEX* is read worldwide and that our friends abroad are contributing to it. Not that long ago, it was impossible for some of them to even get the magazine. Now, they can read about what we are doing and vice versa. That is neat.

The idea of adaptive algorithms for software radios is attractive. We predict that the general idea will be as popular in the amateur world as it has been in the commercial world —perhaps more so. We have a standing invitation to do things that the commercial folks cannot do yet. Let's take advantage of it, eh?

The whole issue of adaptive signal identification represents a very fertile field for experimentation. How can we ignore it? Check out the article in this issue and see for yourself.

In This Issue

Jim Scarlett, KD7O, describes transmit functions in part 3 of the series about his software defined radio. Also, L. B. Cebik, W4RNL, tells us how to make reality converge with design goals in large multiband quad antenna arrays.

Evidently, *QEX* is reaching Eastern Europe, as well as China and other formerly remote areas! We have an article from Oleg Skydan, UR3IQO, about fractional-N synthesizers. Oleg is in Ukraine. He has lately been able to obtain parts to build his unit, the performance of which is outstanding.

Frank Brickle, Jr, AB2KT, offers his thoughts on how to accomplish automatic signal identification in software radios. In our view, this is an exciting area for experimentation for hams. Combined with antenna beamforming and other adaptive techniques, we think there is much promise for future development. John Gibbs, KC7YXD, finishes his series on D-STAR. The system fulfills many of the wishes expressed in response to the ARRL's Technical Working Group (TWG) survey conducted several years ago.

Richard Kiefer, KØDK, brings us an EMI finder that incorporates some useful design techniques at UHF. Even though the design is not centered on a ham band, it does involve frequencies that are useful in looking for interference sources.

In *RF*, Zack Lau, W1VT, tells us about building microwave quads and Yagis.—*Doug Smith, KF6DX;* **kf6dx@arrl.org**. □□

A High-Performance Digital Transceiver Design, Part 3

Transmit functions take center stage.

By Jim Scarlett, KD7O

In Part 1, we discussed an architecture that achieves very good transceiver performance, while entering the digital domain directly from RF.[1] Part 2 showed front-end circuits for the receiver section.[2] In Part 3, we'll follow the transmit path from the input of the Transmit Signal Processor (TSP) to the power amplifier (PA) filter output.

Digital to Analog

The TSP circuit (Fig 1), much like its receive counterpart described in Part 2, is not terribly complex. The most difficult part of this circuit is the package itself. Since it is a quad TSP, there is a higher pin count (128 pins), and the pins are more closely spaced (20 versus 25 mils). This makes soldering by hand something of a challenge. However, while tedious, it's not too bad. Once the corners are tacked down, it's mainly an exercise in patience.

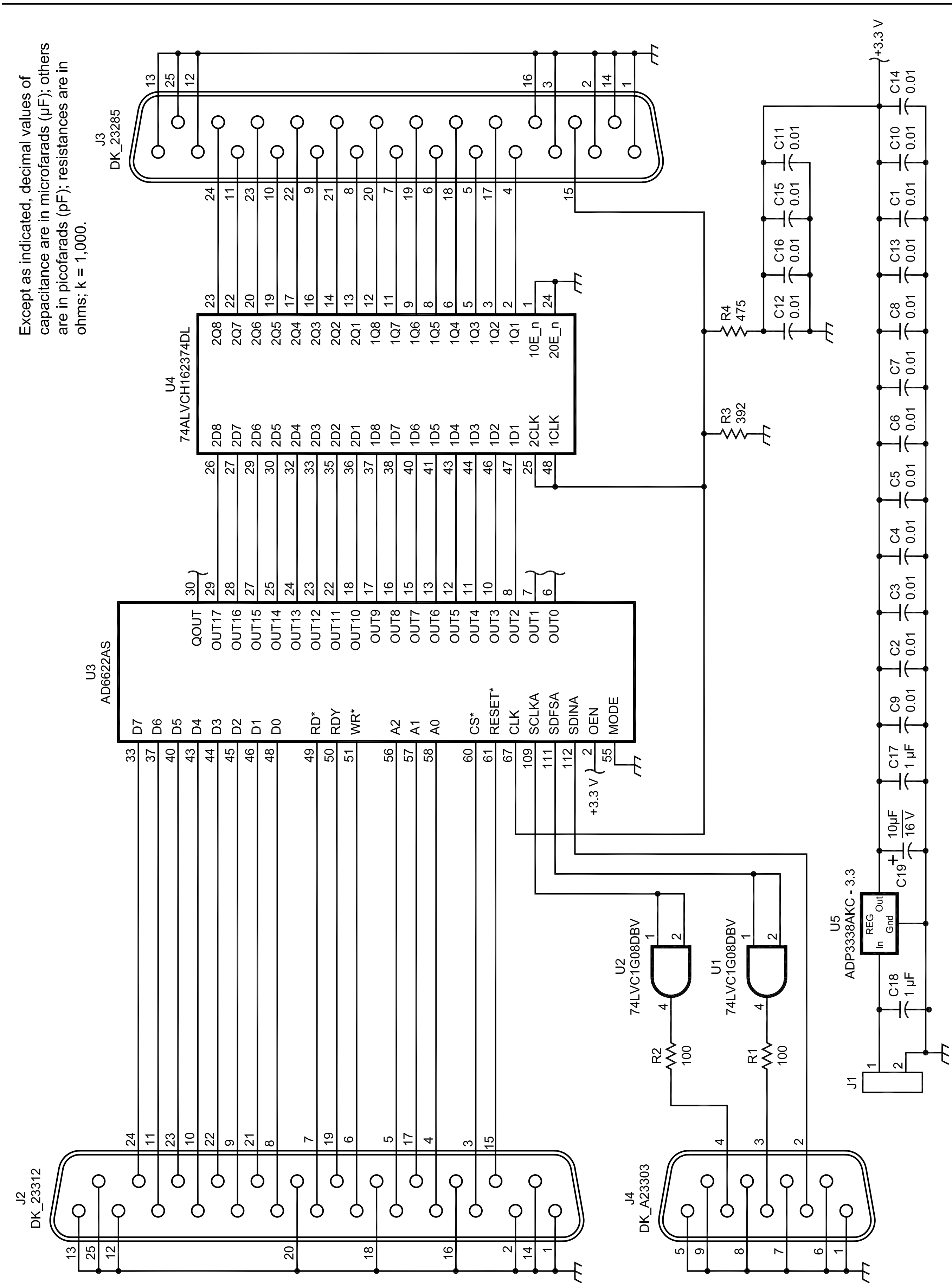
The control interface and serial port are essentially identical to the AD6620 circuit and connect to the board on a DB25 and DB9 connector, respectively. The Microport mode is set to mode 0, so the part can be programmed in the same way as the receive signal processor (RSP). Also, even though the TSP receives serial data from the DSP, it must be the serial master. Therefore, the serial clock and frame sync for channel A are buffered, terminated and sent to the DSP serial port. The clock rate need not be very high, since we will be using 16 bit I/Q data at sampling rates of 16 kHz and 40 kHz.

The TSP is operated at a clock rate of 70.4 MHz. This was determined by several factors. First, it is below the rated value of 75 MHz. This value allows integer interpolation of both 16 kHz and 40 kHz sampling rates. Also, the 64.96 MHz rate used in the receiver would cause the interpolation filter in the transmit DAC to cut off part of the 10 meter band. With the 70.4 MHz clock rate, the entire band is passed.

I was unable to pass the output of the TSP through a simple buffer, like I did between the ADC and the RSP in part 2. The timing was too difficult between the TSP and the DAC. I decided to latch the data and adjust the arrival time of the clock to ensure that

[1]Notes appear on page 11.

10625 Highstream Dr
Raleigh, NC 27614
jim.scarlett@analog.com



Except as indicated, decimal values of capacitance are in microfarads (µF); others are in picofarads (pF); resistances are in ohms; k = 1,000.
J3 DK_23285
U4 74ALVCH162374DL
U3 AD6622AS
J2 DK_23312
J4 DK_A23303
U2 74LVC1G08DBV
U1 74LVC1G08DBV
R1 100
R2 100
R3 392
R4 475
U5 ADP3338AKC - 3.3
REG In Gnd Out
C18 1 µF
C19 10µF 16 V
C17 1 µF
J1
+3.3 V

Fig 1—Transmit signal processor schematic diagram. Resistors are 0805 SMT unless otherwise noted. Capacitors are 0603 SMT unless otherwise noted.
C17, C18—1 µF 16 V X7R 1206 SMT.
C19—10 µF 16 V tantalum.
J1—2 pin header.
J2—DB25 right angle, PC mount (Digi-Key #A23312).
J3—DB25 right angle, PC mount (Digi-Key #A23285.
J4—DB9 right angle, PC mount (Digi-Key (#A23303).
U1, U2—Single AND gate, 74LVC1G08DBV.
U3—TSP, AD6622AS.
U4—16-bit latch, 74ALVCH162374DL.
U5—LDO, ADP3338AKC-3.3.

the required setup and hold times were met.

I chose a 16-bit latch chip that has fast, well-defined timing parameters: the 74ALVCH162374. This part also has series termination resistors built in. The timing is straightforward between the TSP and the latch. The 70.4 MHz clock has a period of approximately 14.2 ns. The '374 is specified to latch correctly if the data is available at least 1.9 ns before the clock (setup time) and if the data remains stable for at least 0.5 ns after the clock (hold time). With this information, we know that new data from the TSP must be valid less than 12.3 ns after each clock, but that it must hold the old data at least 0.5 ns after the clock. From the specification sheet, the part will hold the old data for a minimum of 4.1 ns, and new data is valid in a maximum of 12 ns. Actual timing is likely to be somewhere between these values, but the worst case still meets our timing requirements. Timing between the latch and the DAC will be discussed shortly.

The DAC (Fig 2) is operated on a 140.8 MHz clock to take advantage of its interpolation features. A 70.4 MHz clock could have been used, but would result in more phase noise being superimposed on the transmitted signal.

Three separate 3.3 V supplies are used on the DAC board. For best performance, the datasheet recommends keeping the analog, digital and clock supplies separate, using chokes or some other means.[3] I felt that the easiest (and most thorough) approach would be to use separate regulators for each.

The clock routing circuitry is quite simple. The DAC works best with differential clock input, so a 1:4 transformer is used to generate the differential signal. A dc level shift is applied to the center tap to provide the appropriate offset of Vcc/2. The center tap is also at ac ground. The resistor across the differential signals provides balance and a proper load for the clock driver. A +1 dBm clock input will provide about 1.4 V(pk-pk) to the DAC, which will provide excellent noise performance. A +7 dBm clock input with a 1:1 transformer would do the same, but I had a 1:4 available. For best noise performance, the clock at the DAC needs to be at least 0.5 V(pk-pk).

The internal PLL is disabled by grounding the U2 PLVDD pin. A 70.4-MHz clock is now available at the PLLLOCK pin. This pin has a "fanout" of one, and therefore must be buffered. The 70.4-MHz signal is routed to the TSP and its buffer; the timing relationships involving this clock determine whether the TSP and the DAC play well together.

The period of the 70.4-MHz clock is 14.2 ns. Because of the delay involved in the generation of this clock from the 140.8 MHz input, the setup time is –1.2 ns. This means that the incoming data can arrive up to 1.2 ns *after* the 140.8-MHz rising edge and still be valid. This gives a total available time of 15.4 ns. The maximum delay from the clock input to 70.4 MHz output is 3.2 ns. The maximum propagation delays for the AND gate buffer and the TSP buffer are 4.5 ns and 4.6 ns. Thus, the maximum time for the arrival of the data is 12.3 ns, which provides a margin of 3.1 ns.

The minimum hold time for the DAC is 3.2 ns after the input clock rising edge. The minimum delay from the clock input to 70.4-MHz output is 2.8 ns. The minimum propagation delays for the AND gate buffer and the TSP buffer are 0.8 ns and 1.0 ns. The total minimum propagation delay is 4.6 ns, which provides a margin of 1.4 ns. Both the setup and hold times are met for the interface. Any substitutions of logic should take this timing into account.

Pins 17 and 18 (MOD0 and MOD1) set the operating mode of the device. When pin 17 is low, the interpolation filter is in the low-pass mode. This mode is set for the HF bands. By setting pin 17 high, the filter is in the high-pass mode, which is useful for adapting the transceiver to 6 m as well. Pin 18 determines whether the part is in "zero-stuffing" mode. This mode essentially performs another 2× upsample (with no interpolation filter), which helps to flatten out the $\sin(x)/x$ response of the DAC. The cost is that the maximum output level at lower frequencies will be 6 dB lower. Zero-stuffing is not useful for HF or 6 m, as the signal strength would decrease. I included access to this pin to allow 2-m operation if I decide in the future to include that band.

The DAC outputs are current drivers, with the full-scale current set by a resistor on pin 40 (FSADJ). In this case, the 2-kΩ resistor sets the full-scale current at each output to just under 20 mA. The outputs feed a 1:1 transformer with a center tap. The tap is necessary to provide a dc ground, but it also allows a 6-dB voltage gain. With a 50-Ω load on the transformer output, the peak voltage at either output pin is about 250 mV. This is well within the voltage compliance range and provides some margin below the maximum output level recommended to minimize distortion.

As mentioned above, the output level from the DAC is not constant over the frequency range it will cover. This results from the $\sin(x)/x$ response of DACs. Because of the interpolation in the DAC chip, the roll-off is not as steep as it would be without interpolation. The roll-off is not too bad in the HF range, with the maximum attenuation being less than 0.7 dB at the top of the 10-m band. At 6 m (when added to the radio), the attenuation is still only about 2.2 dB.

Boosting the Power

Ignoring the $\sin(x)/x$ curve, the output of the DAC is about +4 dBm. This is fed into an emitter-follower buffer stage that runs with enough current (about 30 mA) to ensure low distortion. At the output of the buffer is a 9-dB attenuator. The 47.5-Ω resistor presents a good match for a 3-dB attenuator, since the amplifier impedance is essentially r_e, which is very low. The resistor also creates a voltage divider that provides the other 6 dB of attenuation. The attenuator performs two functions. It reduces the signal level so the next amplifier stage operates in a low-distortion mode, and it helps ensure that the band-pass filters are properly terminated.

The buffered DAC output must be filtered to prevent spurious outputs in general, and aliasing in particular. This is one of the primary benefits of interpolation in the DAC. The anti-alias filtering requirements are greatly relaxed, since one-half the sampling rate (the Nyquist bandwidth) is now 70.4 MHz instead of 35.2 MHz. The on-board digital interpolation filter takes care of most images between 35.2 MHz and 70.4 MHz, but images of 10-m signals have far less attenuation. There is some reduction; after all, the 70.4 MHz frequency was used (instead of 64.96 MHz on receive) to put 10 m in the passband of the interpolation filter.

The analog anti-alias filtering is

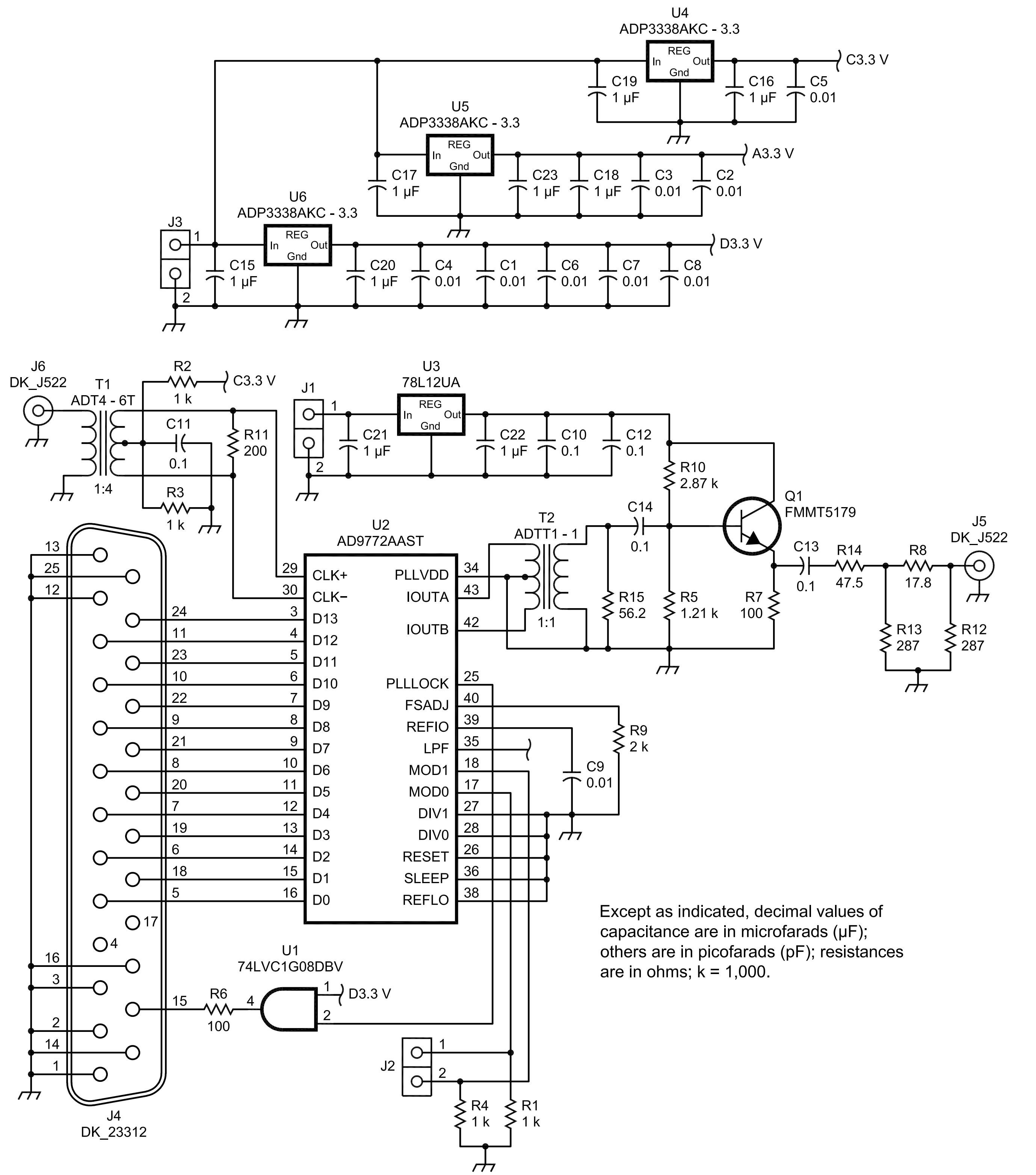


Fig 2—DAC schematic diagram. Resistors are 0805 SMT unless otherwise noted. Capacitors are 0603 SMT unless otherwise noted.

C10—0.01 µF X7R 0805 SMT.
C11-C14—0.1µF X7R 0805 SMT.
C15-C23—1 µF 16 V X7R 1206 SMT.
J1-J3—2 pin header.
J4—DB25 right angle PC mount (Digi-Key #A23312).
J5, J6—PC mount SMB bulkhead jack (Digi-Key #J522).
Q1—RF NPN transistor FMMT5179.
T1—1:4 transformer, CT (Minicircuits ADT4-6T).
U1—Single AND gate, 74LVC1G08DBV.
U2—High-speed DAC, AD9772AAST.
U3, U4—Low power 12 V regulator, 78L12UA.
U5, U6—LDO, ADP3338AKC-3.3.

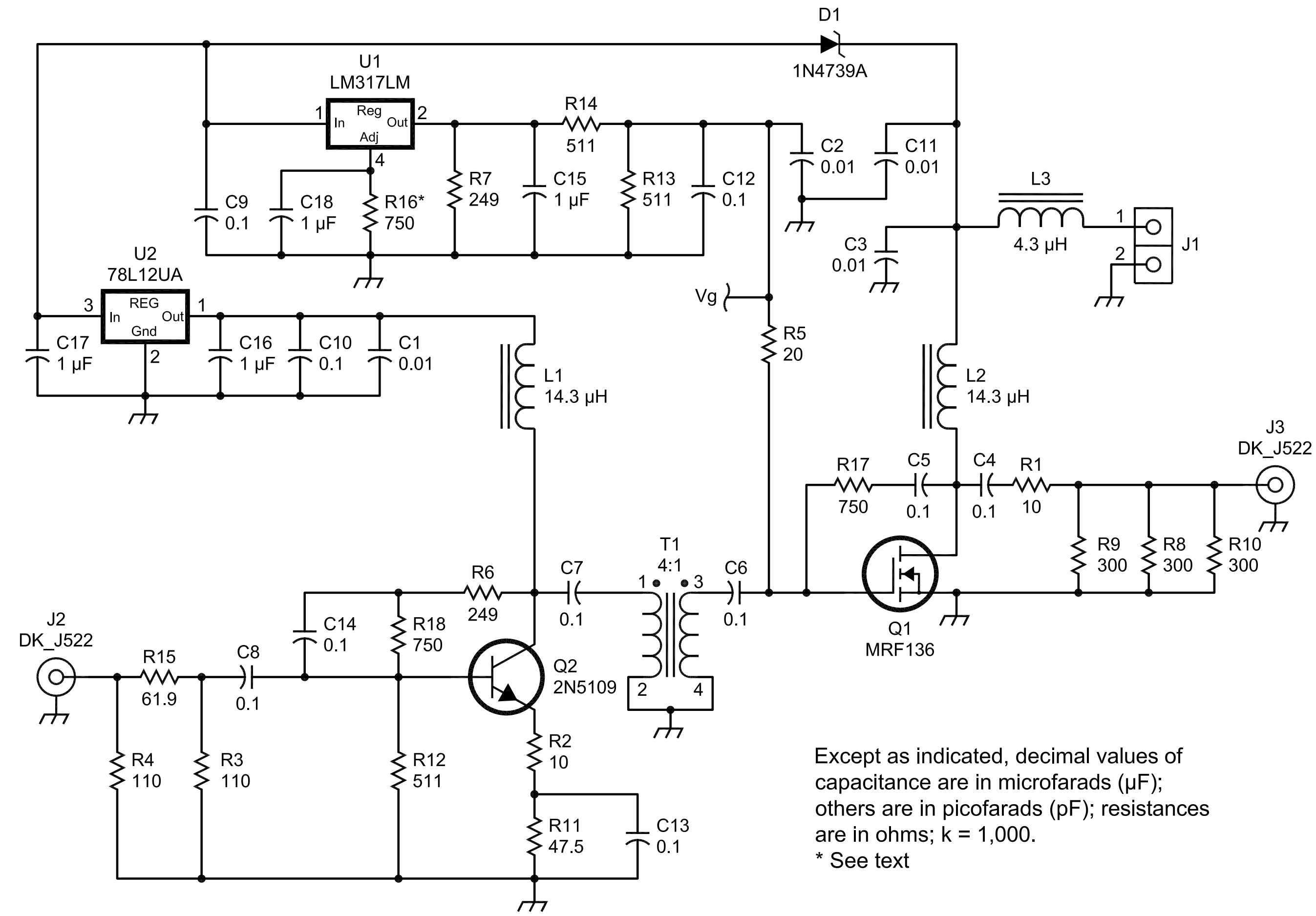


Fig 3—Driver schematic diagram. Resistors and capacitors are 0805 SMT unless otherwise noted.

C15-C18—1 µF 25 V X7R 1206 SMT.
D1—9.1 V 1 W Zener, 1N4739A.
J1—2 pin header.
J2, J3—PC mount SMB bulkhead jack (Digi-Key #J522).
L1, L2—14 t #22 AWG on a FT61-50A core.
L3—8 t #22 AWG on a FT61-50 core.
Q1—MRF136 MOSFET.
Q2—RF NPN transistor, 2N5109.
R8-R10—300 Ω 1/4 W 1206 SMT.
R11—47.5 Ω 1/4 W 1206 SMT.
R16—750 Ω test select for 300 mA drain current in Q1.
T1—BN202-43 core: secondary, 1 t (1/8-inch brass tubing), primary, 2 t #22 AWG through secondary.

done by the same filter/amplifier chain that is used on receive. On the transmit side, both amplifiers are used, to get approximately 17 dB of gain. The filtering in this chain, which was necessary to meet stringent receiver requirements, is more than enough for this application. The output is about +12 dBm.

A two-stage driver follows the filter/amplifier chain (Fig 3). One stage was not enough to reach the driver output target of 500 mW. Therefore, a two-stage design was used, with a 9 dB attenuator at the input. The attenuator keeps the amplifiers from being overdriven and provides an excellent termination for the filters. The first stage is a common-emitter BJT amplifier using emitter degeneration and shunt feedback. There is nothing unusual about this amplifier, as readers of *Solid State Design* can readily tell.[4] The first stage provides about 10 dB of gain.

The second stage of the driver is a common-source amplifier using an MRF136 FET. This device is capable of excellent performance well into the VHF range.[5] The amplifier was designed with possible expansion to 6 m in mind.

The 20-Ω resistor on the gate of the FET helps ensure stability, and also makes the input easier to match across a broad range of frequencies. A 4:1 transformer completes the input network. The feedback resistor is there more to help achieve good input and output matches than for stability, which should be assured by the gate resistor. The output network consists of only a series and shunt resistor. An L-network could be added for an even better return loss, but it is not necessary for driving the PA. The driver is mounted on the same heat sink as the PA.

Gate bias is provided via a low-power adjustable regulator. The resistance from the regulator to ground is test selected for a drain current in the MRF136 of approximately 300 mA. This is enough current to achieve excellent linearity in class-A service. The maximum output required from the driver is about 500 mW, though it is capable of much more. I think enough of the MRF136 that I'm considering replacing the receiver preamplifiers with these devices running with high currents.

Power Amplifier

The power amplifier uses a single MRF151 FET to generate a little over 60 W PEP or CW, and runs on 40 V (Fig 4). The device is capable of considerably more, but I didn't feel that I

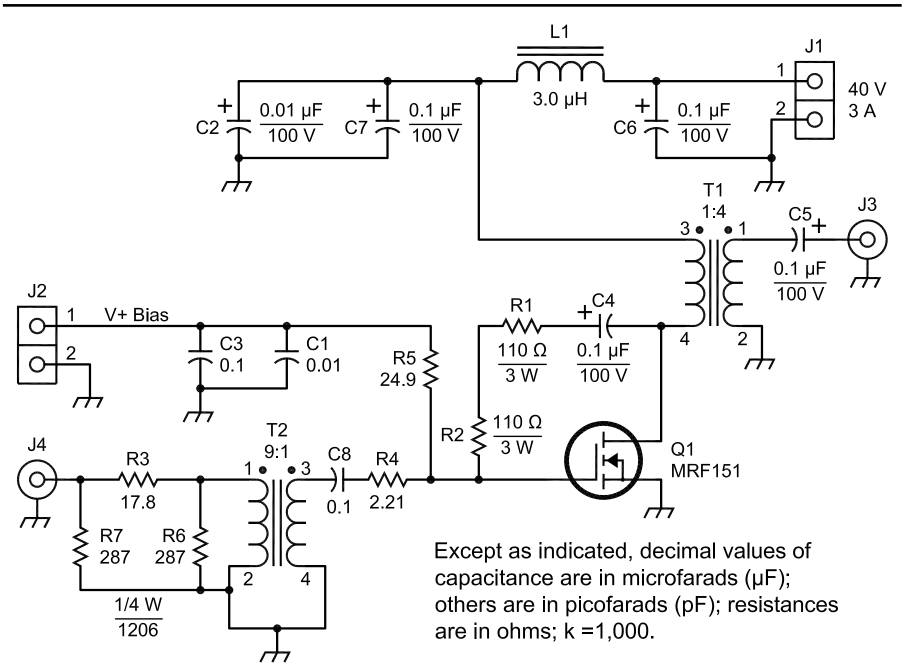


Fig 4—Power amplifier schematic diagram. Resistors and capacitors are 0805 SMT unless otherwise noted.
C2—0.01 μF 100 V.
C4-C7—0.1μF 100 V.
J1, J2—2 pin header.
J3, J4—BNC jack.
L1—16 t #18 AWG on a T106-6 core.
Q1—MRF151 MOSFET.
R1, R2—110 Ω, 3 W metal oxide.
R3—17.8 Ω ¼ W 1206 SMT.
R6, R7—287 Ω ¼ W 1206 SMT.
T1—BN7051-43 core: primary, 1 t (¼-inch brass tubing); secondary, 2 t #18 AWG through primary.
T2—BN3312-43 core: secondary, 1 t (3/16-inch brass tubing); primary, 3 t #22 AWG through secondary.

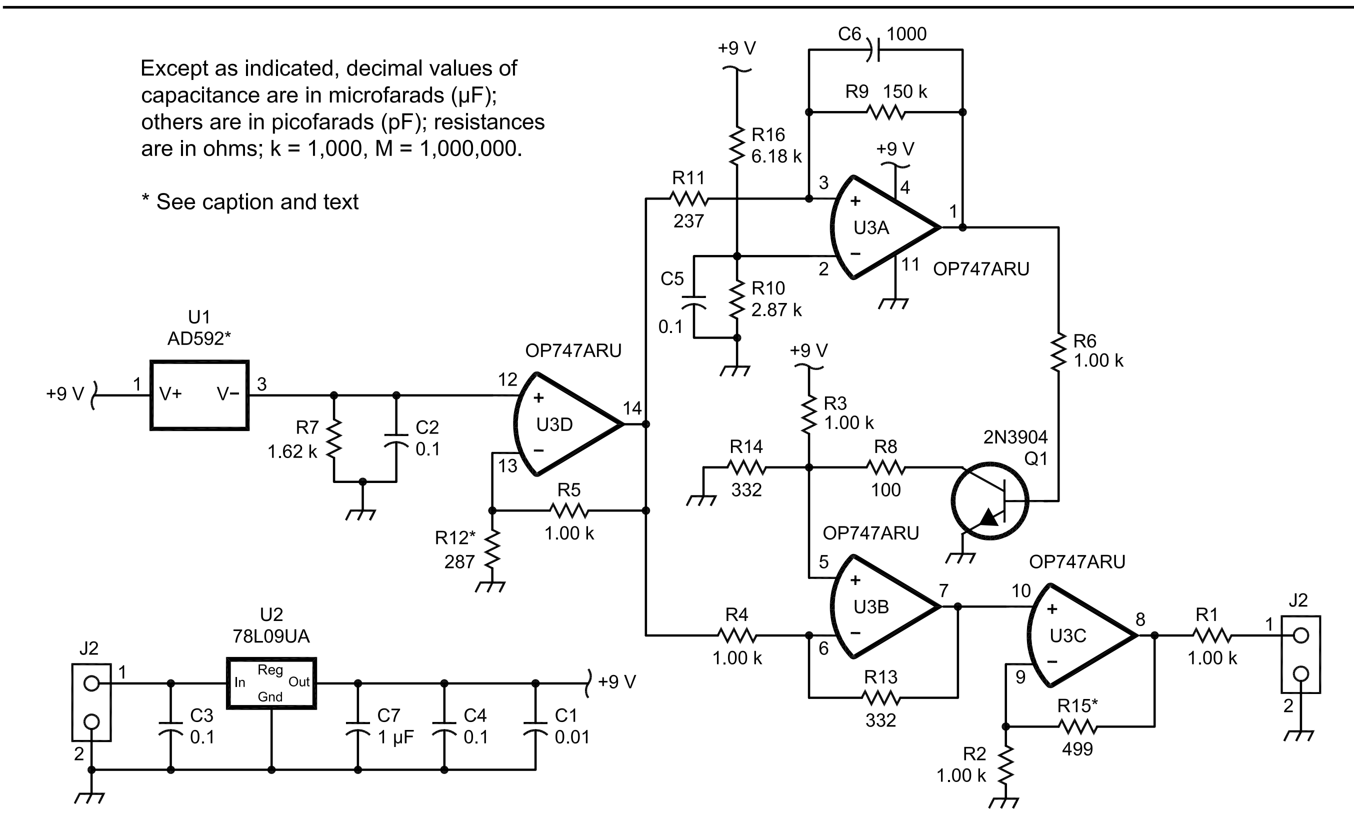


Fig 5—Power amplifier bias schematic diagram. Resistors and capacitors are 0805 SMT unless otherwise noted.

U1—AD592 temperature transducer. Mount with thermal compound on MRF 151.
U2—78L09μA +9 V voltage regulator.
U3—Quad opamp, OP747ARU.
C7—1 μF, 25 V X7R 1206 SMT.
J1, J2—2 pin header.
Q1—NPN transistor, 2N3904.
R10—2.87 kΩ, 0805 SMT. Can be adjusted to set bias shutdown threshold.
R12—287 Ω, 0805 SMT. Test select for stable bias over temperature.
R15—499 Ω, 0805 SMT. Test select for proper bias level (IDQ = 500 mA).

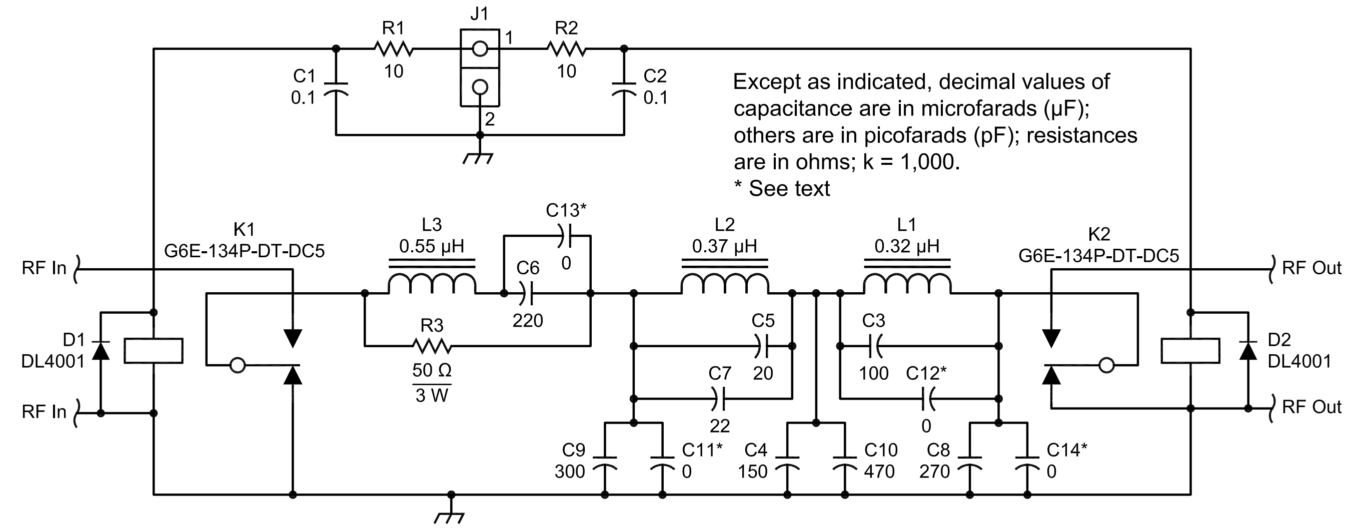


Fig 6—Transmit low-pass filter schematic diagram. Values shown are for the 20 meter filters. Component values for the other HF bands are available on the ARRLWeb (see Note 9). Resistors are 0805 SMT unless otherwise noted. Capacitors are Silver Mica unless otherwise noted.

0 pF capacitors are pads for Silver Mica capacitors.
C1, C2—0.1 µF X7R 0805 SMT.
D1, D2—DL4001.
J1—2 pin header.
K1, K2—PC mount SPDT relay (Digi-Key # G6E-134P-ST-US-DC5).
L1—0.32 µH, 13 t #18 AWG on a T106-0 core.
L2—0.37 µH, 14 t #18 AWG on a T106-0 core.
L3—0.55 µH, 17t #18 AWG on a T106-0 core.

really needed it. This power level is perfectly adequate for "barefoot" operation, and it does not tax the devices involved terribly much. It also allowed me to design the power supply mostly with parts I had on hand. I also have some plans for a 300-400 W amplifier, which would only require about 20 W of drive.

The circuit for the amplifier has a lot of similarities to that in the MRF151 datasheet.[6] A 9:1 transformer and series resistor match the driver output with the gate of the device. I chose to use a conventional transformer with a one-turn secondary consisting of brass tubing, because of its easy construction. Like the driver circuit, a resistor is placed from gate to ground to tame the beast and broaden the input response. For this device, the recommended resistance is 25 Ω. The feedback resistance helps create decent input and output matches, and flattens the gain across the spectrum. A 1:4 output transformer provides a load line optimized for 64 W output. Again, I opted for the conventional transformer, this time with a larger core. More heat will be generated than with a transmission-line transformer (see the discussion in Note 8), but this configuration will work fine.

The bias network (Fig 5) requires some explanation. In reading about other FET amplifier projects, I noticed a couple of different modes of thermal protection. In his 50-MHz amplifier, Dick Frey, K4XU, used thermal compensation to account for the fact that as the FET heats up, the same gate drive will induce larger standing currents (thus generating more heat).[7] Bill Sabin, WØIYH, used thermal monitoring to shut down the devices if the temperature reached a certain threshold.[8]

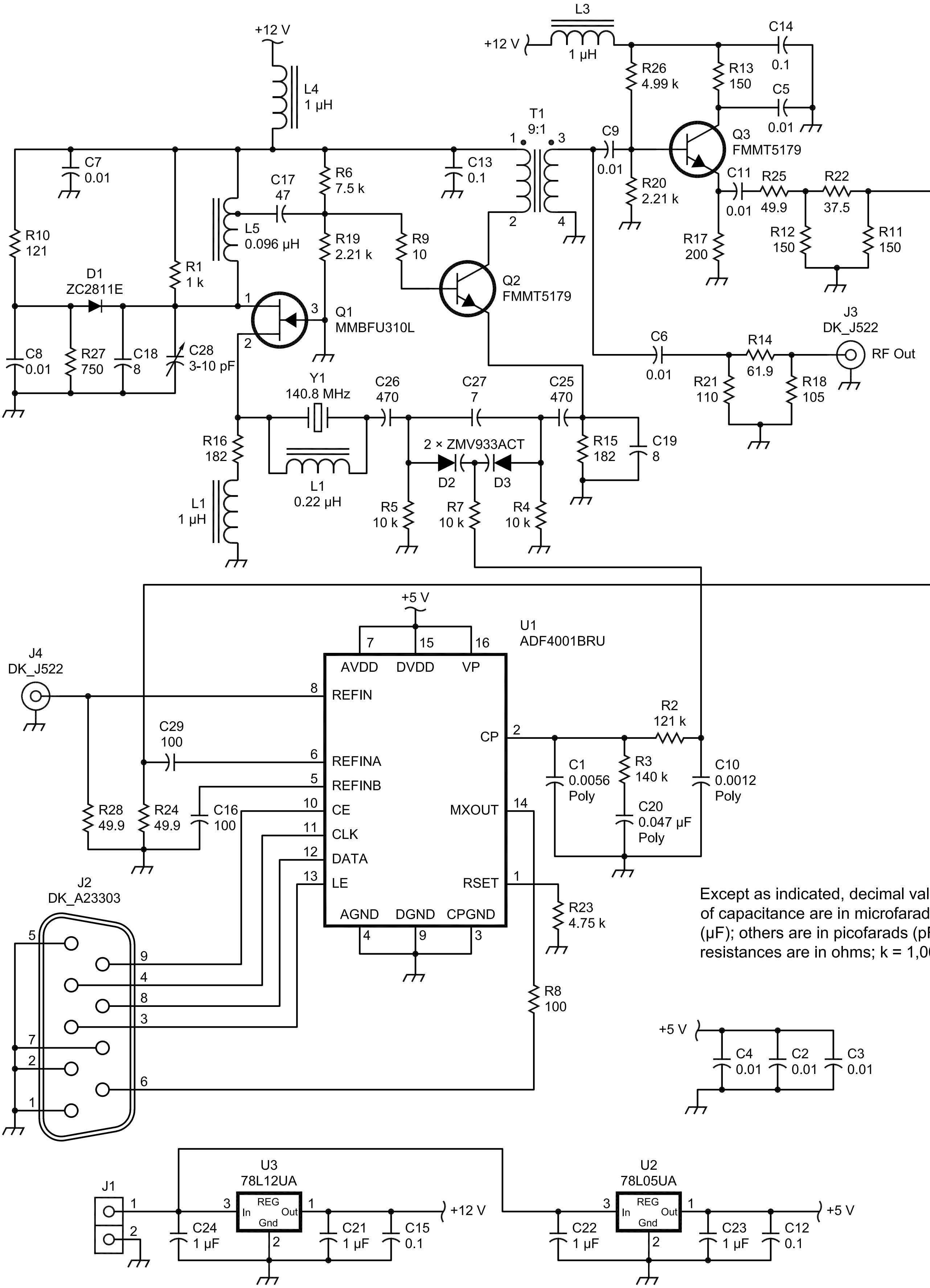
I decided that I liked the idea of doing both functions, though I took a slightly different approach. Accomplishing both functions with a single temperature sensor requires a little more complexity than a single function. I used a current-output temperature sensor, the AD592, that has a more linear response than thermistors. The output of this device is 1 µA per Kelvin (at +25°C, the output is 298 µA). The TO-92 package is mounted directly on the MRF151.

The output of the AD592 is fed to a buffer amplifier whose input uses a 1.62-kΩ resistor to convert the current signal to a voltage. The buffer has a gain of about 4.5, which sets the slope of the temperature compensation. The value of R12 can be adjusted to ensure that the bias remains constant over temperature. The buffer feeds two other amplifiers. One is a summing amplifier that combines the bias voltage with the compensation voltage. The other is used as a comparator to clamp the summing amplifier output to about 0 V when the MRF151 temperature gets too high. The hysteresis in the comparator circuit creates a window of about 1.9°C, meaning that the temperature must drop that amount before bias is restored. The feedback capacitor changes the hysteresis for high-frequency signals, so that false triggering from this source does not occur.

The summing amplifier feeds a fourth amplifier that sets the final output. The value of R15 sets the bias point. This can be set while observing the two-tone output on a spectrum analyzer; or if one is not available, to set the I_{DQ} to about 500 mA. This last amplifier is not absolutely necessary, but I wanted to make it as simple as possible to adjust the bias. If instead the bias were to be adjusted using R13 in the summing amplifier, the value of R14 must be adjusted also. Otherwise, the compensation slope would be affected by the gain change. Besides, I still had a fourth amplifier available in the package. The calculations used for the bias network are in a spreadsheet ("PA_bias_compensation.xls") that can be downloaded from the ARRLWeb.[9]

Low-Pass Filters

The output low-pass filters (Fig 6) are of the Cauer type. They have been optimized for suppression of the sec-



Except as indicated, decimal values of capacitance are in microfarads (µF); others are in picofarads (pF); resistances are in ohms; k = 1,000.

Fig 7—Transmit PLL schematic diagram. Resistors and capacitors are 0805 SMT unless otherwise noted.
C1—0.0056 µF polyester (Digi-Key #P3562).
C2-4—0.01 µF 0603 SMT.
C10—0.0012 µF polyester (Digi-Key #P3122).
C20—0.047 µF polyester (Digi-Key #P3473).
C21, C24—1 µF 25 V X7R 1206 SMT.
C22, C23—1 µF 16 V X7R 1206 SMT.
C28—3-10 pF trimmer, SMT (Digi-Key #SG2002).
D1—Schottky diode, ZC2811E.
D2, D3—Tuning diode (Digi-Key #ZMV933ACT).
J1—2 pin header.
J2—DB9 right-angle PC mount (Digi-Key #A23303).
J3, J4—PC mount SMB bulkhead jack (Digi-Key #J522).
L1—0.22 µH.
L2-L4—1.0 µH.
L5—8 t #28 AWG on a T37-12 core; tap 2 t from "cold" end.
T1—primary 6 t #28 AWG on a BN2402-61 core; secondary 2 t #28 AWG.
U1—PLL, ADF4001BRU.
U2—regulator, SOT89, 78L05UA.
U3—regulator, SOT89, 78L12UA.
Y1—140.8 MHz, seventh overtone (International Crystal).

ond and third harmonics, as described by Tonne.[10] Also, they incorporate a diplexer as Stephensen demonstrated to provide a good match for the harmonics. This reduces reflections of the harmonics and improves the linearity of the amplifier output.[11]

The iron-powder cores used in the filters were chosen to ensure that overheating would not be a problem at this power level. The objective was to keep the flux density for each inductor significantly below the maximum recommendations outlined in the Amidon data manual.[12] Accomplishing this involves larger cores and more turns, which usually don't go hand in hand. This is why, for instance, the 20-m filter inductors use the "0" material. The capacitors are the silver-mica types, with a rating of 500 V. They can easily handle the signal levels expected in the filters.

Transmit Clock

The design of the transmit clock is nearly identical to the low-noise clock used for receiving (Fig 1). Several component values have been changed to reflect the much higher frequency of operation. Also, remembering that we only need +1 dBm for the DAC clock, the output has additional attenuation.

The PLL reference frequency is 64 kHz. This is necessary because the reference oscillator frequency was selected to minimize noise in the receive PLL. Still, the final result works out pretty well. The loop bandwidth is only about 65 Hz, because the PLL noise is much higher with the large divide ratio ($N = 2200$). Inside the loop, the PLL noise dominates (the reference noise is much lower).

Outside the loop, the VCXO noise dominates. The predicted phase noise of the VCXO is quite good; though as Leeson predicts, it is not as good as the 64.96-MHz receive oscillator. However, the final noise performance on the transmit side is not considerably different than for receive. This is because of the effect that phase noise (or jitter) has on ADCs and DACs. As shown in the "Phase Noise and ADC Performance" sidebar in Part 2 (see Note 2), the effect is smaller for lower signal frequencies. So, the phase noise of the transmit clock, when applied to a 10-m signal, is reduced by about 14 dB. The noise sidebands actually applied to the desired signal in both transmit and receive are therefore quite similar.

Summary

Modern signal-processing devices can give us as much simplification in our transmitter architecture as on receive architecture. In this design, we have barely scratched the surface of what can be done with all of the flexibility built into the TSP chip. The performance of the DAC gives us a very low-noise, low-distortion output capability. DAC features, such as interpolation, give us much more flexibility in the analog design.

New devices, either released or about to be released, are improving on even this excellent performance. They provide even more flexibility than that found in the AD9772A. For that reason, this design was done in a modular fashion, just as was the receiver. For instance, a new 16-bit DAC design can be directly interfaced to the TSP output connector to upgrade performance. As such, we can constantly improve the performance of this radio as new technology allows. Additionally, we can add new modes simply by upgrading the software or changing the configuration of the TSP.

We have now looked at the main signal-processing blocks for both receive and transmit. Next time, we will look at some of the circuits that bring it all together, including audio and control blocks. The DSP will be linked into the system as well.

I would like to thank all the gentlemen whose work is referenced in this article. Their excellent work made mine much easier.

Notes

[1] J. Scarlett, KD7O, "A High-Performance Digital Transceiver Design, Part 1", *QEX*, Jul/Aug 2002, pp 35-44.
[2] J. Scarlett,KD7O, "A High-Performance Digital Transceiver Design, Part 2", *QEX*, Mar/Apr 2003, pp 3-12.
[3] Analog Devices, AD9772A datasheet, Rev A, 3/02.
[4] D. DeMaw, W1FB, and W Hayward, W7ZOI, *Solid State Design for the Radio Amateur*, (Newington, Connecticut: ARRL, 1977), pp 188-192.
[5] M/A-Com, MRF136 datasheet, Rev 7.
[6] M/A-Com, MRF151 datasheet, Rev 9.
[7] R. Frey, K4XU, "A 300-W MOSFET Linear Amplifier for 50 MHz," *QEX*, May/Jun 1999, pp 50-54.
[8] W. Sabin, WØIYH, "A 100-W MOSFET HF Amplifier," *QEX*, Nov/Dec 1999, pp 31-40.
[9] You can download this package from the ARRLWeb at **www.arrl.org/qexfiles/**. Look for 1103SCARLETT.ZIP.
[10] J. Tonne, WB6BLD, "Harmonic Filters, Improved," *QEX*, Sep/Oct 1998, pp 50-53.
[11] J. Stephensen, KD6OZH, "The ATR-2000: A Homemade, High-Performance HF Transceiver, Part 3," *QEX*, Mar/Apr 2001, pp. 3-8.
[12] Amidon Associates Data Manual, Jan 2000, pp 1.35-1.41.

Notes on Designing Large Five-Band Quads

Would you like to design a large loop Yagi and achieve the design performance in reality? Here's how.

By L. B. Cebik, W4RNL

The design of large five-band quad arrays has a number of facets, each of which deserves attention by the would-be quad user. We can divide them into three general groups:

1. The use of antenna modeling software as the design vehicle: How do we set up the model for effective design work?
2. The performance of the quad as designed: How can we use the modeled performance as a guide to evaluating and improving designs?
3. The transition from model to physical antenna: What factors play a role in determining if and how the modeled array should be built?

1434 High Mesa Dr
Knoxville, TN 37938-4443
e-mail **cebik@cebik.com**

Although it is not possible to provide definitive answers to all of these questions, we can run through a design exercise and extract as much guidance from it as possible. Although not exhaustive, the amount of guidance will be considerable.

For our project, let's consider the design of one or more large five-band HF quad arrays. By large, I mean an array with at least four elements per band.

Setting Up the Design Project

The availability of *NEC*-based antenna modeling programs has moved much of the design process from the antenna tower to the computer. However, the process of design may prove daunting unless we approach it in a somewhat systematic manner.

Constraints

Designing a large quad array involves some concessions to reality from the start. For example, multiband quad arrays typically employ planar groups of elements: that is, flat, four-arm nonconductive structures to support an element for each of the bands of concern. Consequently, the designer cannot, for each band, select the optimal spacing between elements for maximizing key performance parameters, such as gain, front-to-back ratio (F/B) and SWR bandwidth. Every performance outcome will be a compromise, with its foundation in the initial spacing decisions for the sets of support arms.

Equally limiting will be the fact that quad arrays typically use wire elements. At the outset, I shall specify #12 AWG copper wire as the material of choice for this exercise. However, that very choice will limit and direct the design effort. As I have shown elsewhere, the gain and the operating bandwidth (in terms of both F/B and the 2:1 SWR curves) are functions of the element diameter when specified as a fraction of a wavelength.[1] In the upper-HF region, #12 wire is a small fraction of a wavelength. Achieving full operating potential requires element diameters approaching about 0.5 inch at 10 meters and 1 inch at 20 meters.

However, the planar arrangement of elements does permit the quad designer to achieve—at least on some bands—a higher level of performance than would be provided by a monoband version of the array using similar dimensions.[2] The effects of element interactions on the large quad array will be among the phenomena we shall examine.

A Starting Point

Because many examples of large quad-array design already exist, we need not begin at random. One of the better designs available is the product of Danny Mees, ON7NQ.[3] It consists of three elements on 20, 17 and 15 meters, with a fourth element added for 12 and 10 meters. As a three-element quad on the lower three bands, the array uses a familiar set of element spacing. As shown in part of Fig 1, the reflector is 10 feet from the driver, with a director 8 feet forward of the driver. On 12 and 10 meters, Danny added new elements centered between the reflector and the driver. The new element becomes the driver for the upper two bands, with two directors in front of it. Table 1 supplies the modeled dimensions of the ON7NQ 3-4-element quad and the dimensions of the other large quads we shall explore. Fig 2 shows the general outline of the entire ON7NQ array.

Since one facet of quad design is reducing the number of variables involved, we will use the initial spacing selections of the ON7NQ array as a starting point. Then we shall add one or more elements to each band. In Fig 1B you can see that an additional director has been added, once more at the standard 8-foot spacing from the ON7NQ forward element, resulting in a 26-foot boom. Thus we have a 4-5-element array. Fig 3 shows an outline sketch of the full set of elements.

Fig 1C shows the layout using a wider spacing for the new director. This places the forward elements 12 feet from the ON7NQ forward elements, resulting in a 30-foot boom length. However, for reasons that will become clear as we proceed to analyze the design, it became necessary to add another partial element set equally spaced between the two forwardmost full element sets. The new support holds elements only for 15 and 10 meters. The end product is a 4-5-6-element quad array, shown in Fig 4.

[1]Notes appear on page 24.

Specifications for 4-5-Element Quad

The design process could proceed without a set of goals, but then you would never know when to stop. A set of clear specifications, based on reasonable expectations that emerge from experience, can direct the work of optimizing a design. This converts an endless task into a merely long but finite one. For the design project at

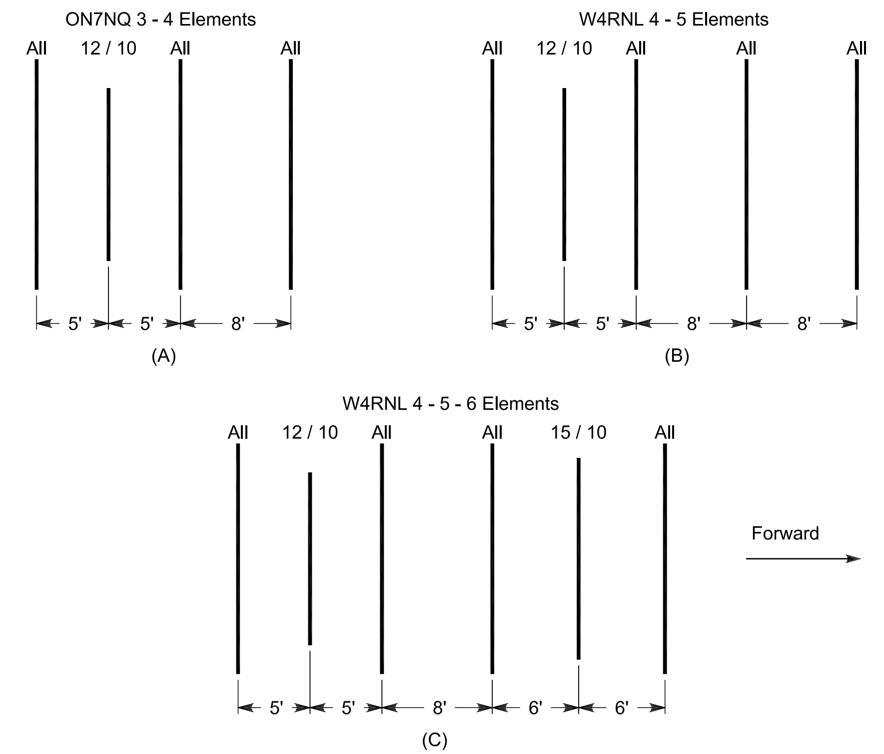


Fig 1—Element spacing for three large five-band quad designs.

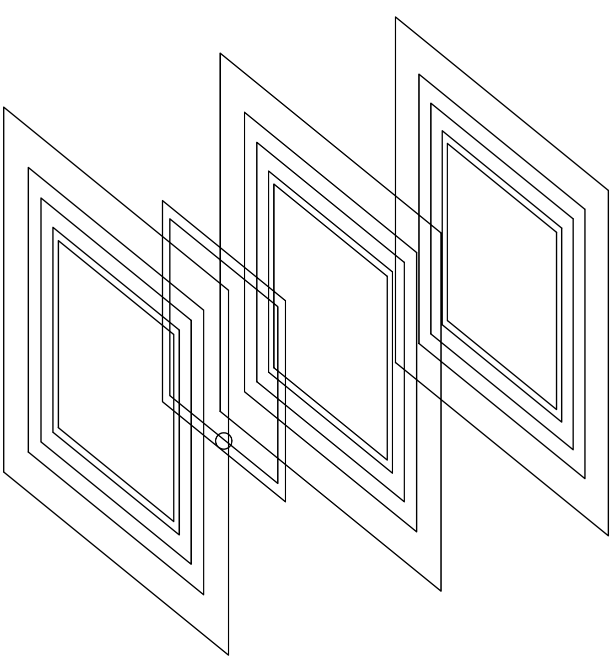


Fig 2—Outline sketch of the 3-4-element ON7NQ five-band quad.

Table 1—Three-Large 5-Band Quad Array Dimensions

ON7NQ, 3-4-Element 5-Band Quad Dimensions (Inches)

Antenna Part	*Side Length*	*Loop Circumference*
20 Refl	217.0	868.0
20 Driver	213.7	854.8
20 Dir 1	205.0	820.0
17 Refl	168.5	674.0
17 Driver	166.3	665.2
17 Dir 1	159.8	639.2
15 Refl	144.8	579.2
15 Driver	142.0	568.0
15 Dir 1	138.0	552.0
12 Refl	122.4	489.6
12 Driver	119.9	479.6
12 Dir 1	118.2	472.8
12 Dir 2	118.7	474.8
10 Refl	110.68	442.7
10 Driver	105.8	423.2
10 Dir 1	104.6	418.4
10 Dir 2	103.99	416.0

W4RNL, 4-5-Element 5-Band Quad Dimensions (Inches)

Antenna Part	*Side Length*	*Loop Circumference*
20 Refl	217.0	868.0
20 Driver	213.0	852.0
20 Dir 1	195.0	780.0
20 Dir 2	196.0	784.0
17 Refl	168.5	674.0
17 Driver	165.6	662.4
17 Dir 1	159.8	639.2
17 Dir 2	159.8	639.2
15 Refl	145.4	581.6
15 Driver	141.4	565.6
15 Dir 1	139.5	558.0
15 Dir 2	139.3	557.2
12 Refl	122.4	489.6
12 Driver	120.6	482.4
12 Dir 1	118.2	472.8
12 Dir 2	119.8	479.2
12 Dir 3	118.6	474.4
10 Refl	110.0	440.0
10 Driver	105.8	423.2
10 Dir 1	104.4	417.6
10 Dir 2	105.0	420.0
10 Dir 3	104.0	416.0

W4RNL, 4-5-6-Element 5-Band Quad Dimensions(Inches)

Antenna Part	*Side Length*	*Loop Circumference*
20 Refl	217.0	868.0
20 Driver	213.0	852.0
20 Dir 1	201.2	804.8
20 Dir 2	194.8	779.2
17 Refl	168.5	674.0
17 Driver	165.6	662.4
17 Dir 1	160.2	640.8
17 Dir 2	159.6	638.4
15 Refl	145.8	583.2
15 Driver	141.4	565.6
15 Dir 1	139.0	556.0
15 Dir 2	138.8	555.2
15 Dir 3	138.4	553.6
12 Refl	121.8	487.2
12 Driver	120.2	480.8
12 Dir 1	119.4	477.6
12 Dir 2	120.2	480.8
12 Dir 3	117.4	469.6
10 Refl	110.6	442.4
10 Driver	105.7	422.8
10 Dir 1	104.4	417.6
10 Dir 2	104.8	419.2
10 Dir 3	104.8	419.2
10 Dir 4	104.2	416.8

See text and Fig 1 for element spacing data.

hand, the following specifications were set for the 4-5-element quad.

Gain: The average free-space gain of the 4-5-element quad should be about 0.7 dB higher than the ON7NQ array on each band. This goal is likely to be achieved on all but 20 meters, where the boom length is short for four elements. The length is adequate for a monoband optimized Yagi, but the fixed spacing of the quad array limits improvements. First, a monoband quad generally requires greater spacing than a monoband Yagi to achieve its full gain potential for any given element diameter. Second, on 20 meters the elements do not have other elements outward from which to potentially derive a modicum of performance enhancement. Third, the individual element spacing may not be optimal. When the spacing was increased to the 30-foot boom length for the 4-5-6-element array, the gain specification was raised by about 0.2 dB as the design goal.

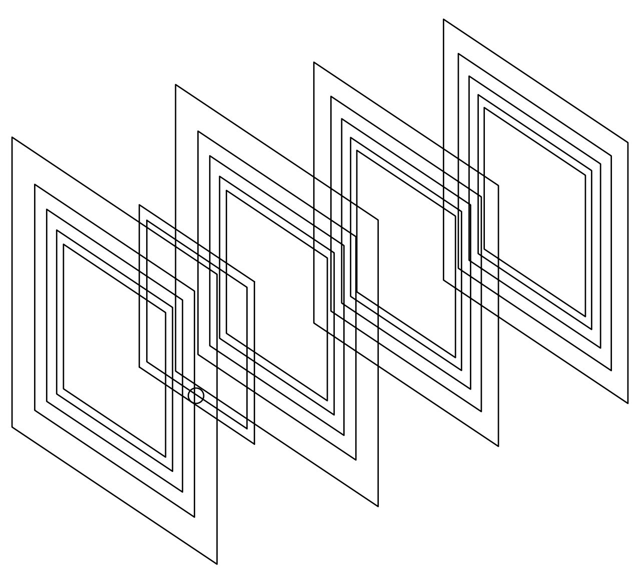


Fig 3—Outline sketch of the 4-5-element W4RNL five-band quad.

Front-to-Back Ratio: It is almost impossible to obtain a 20-dB front-to-back ratio from a wire quad across any HF band (except for the narrow WARC bands). Consequently, the 20-dB standard, long applied to monoband Yagi designs, had to be set aside. More realistic is a goal of achieving a 15 dB front-to-back ratio across the bands. Even this reduced standard cannot be achieved on every band with every configuration. Part of the analysis will deal with why some bands fall short of this goal for some array configurations.

The front-to-back specification is given in terms of the 180° front-to-back ratio. Due to element interactions and the fixed spacing of the elements, a full front-to-rear evaluation will only sometimes match the 180° front-to-back ratio. A front-to-rear evaluation examines the entirety of the radiation pattern to the rear of the beam. Large multiband quad array rear patterns can range from good to exceptionally "messy."

Feed-Point Impedance: Since the ON7NQ array was designed for direct feed with a 50-Ω coaxial cable, the larger arrays also use this feed-point impedance as a specification. The usual 2:1 SWR standard will be applied to determine if the feedpoint impedance falls within the range limits.

Bandwidth Coverage: Although the ON7NQ array was optimized for the CW end of each of the HF bands covered, the goal for the larger arrays was to allow operation over each band. This was not always possible. The 20-meter band is especially resistant to full coverage within the other performance specifications. The 10-meter band was also limited to coverage of the first 800 kHz (from 28.0 to 28.8 MHz), since broader coverage required a severe reduction in performance levels.

Design Strategy

The discussion of a starting point and a set of specifications involve basic "whats," but they do not tell us anything of the "how" of design. Design work with antenna-modeling software requires a strategy if the work is to proceed effectively, even efficiently. Modeling a five-band quad with more than three elements results in a large model.

Moreover, each element that will be modified in the design process involves—assuming a free-space model —the alteration of 16 coordinate values for each and every change. For the task at hand, modeling software that permits the use of variables as coordinates can simplify the work of alteration to a single operation. Consequently, the design work was performed using *NEC-Win Plus*, which permits 24 variable assignments—just enough for the entire project without resorting to workarounds.

Another strategic issue is the segmentation of the element loops. Ideally, a good *NEC* model attempts to align segment junctions to achieve maximum accuracy. Since each wire is to some degree active on all bands, the 20-meter elements should have about twice the number of segments as the 10-meter elements, so that each segment is about the same length at the highest frequency to be used. Since five segments per side is about the minimum level of segmentation to assure accurate results with a closed-loop structure, the overall segmentation becomes a matter of number juggling.

If we place seven segments on each side of a 10-meter element, and if we increase the number by two for each lower band in order, we arrive at 15 segments per side on the 20-meter elements. Fig 5 sketches the elements and the suggested level of segmentation for one set of elements for five bands. This segmentation scheme comes close to meeting the desired 2:1 ratio of segments between 20 and 10 meters and yields a practical alignment of segment junctions from one band to the next.

The resulting arrays are large in both the number of wires and the number of segments. A fully segmented ON7NQ array requires 68 wires and 724 segments, already more sizable than the limits of some widely used software. The 4-5-element array needs 88 wires and 944 segments, while the final 4-5-6-element quad

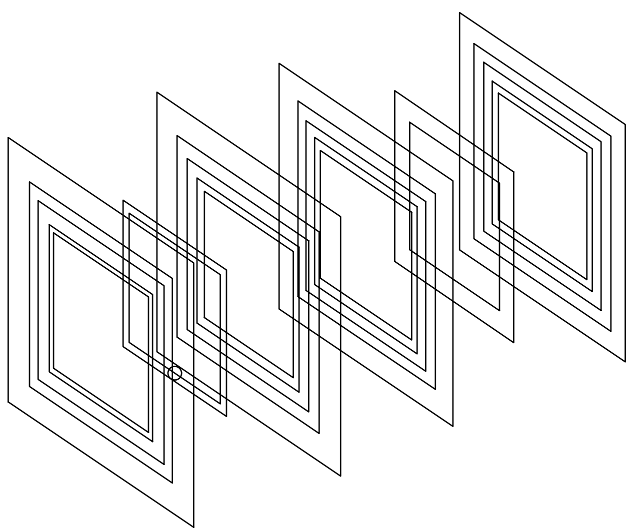


Fig 4—Outline sketch of the 4-5-6- element W4RNL five-band quad.

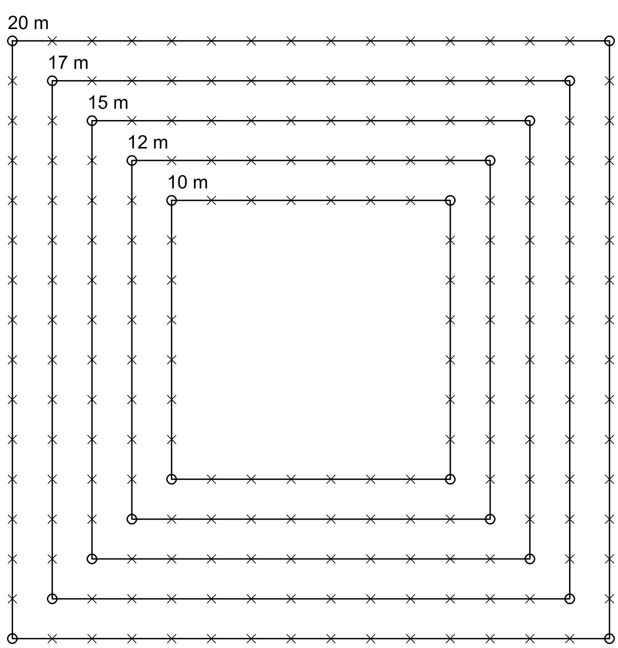


Fig 5—Recommended "full" segmentation of a five-band set of quad elements. Notice how the segments align among the elements.

calls for 96 wires and 1016 segments. Since the time required for each run of the model goes up by the power of the number of wires and the number of segments, core runs for the largest array can approach the limit of most people's patience!

The solution is to use a lower level of segmentation, but only after verifying its adequacy, if not its accuracy. Therefore, I ran a set of comparative curves on the ON7NQ array using full segmentation and a lower level: five segments per side for the upper three bands, and seven segments per side for the lowest two bands. The resulting models produced operating bandwidth data such that in only two instances was a final small adjustment required using the fully segmented model. However, the actual gain and front-to-back figures were sufficiently off that only the performance trends were used to optimize the model. The final tables reflect the performance for the fully segmented models.

Table 1 thus reflects the dimensions for the fully segmented arrays. Without the preparation outlined above, the few hours of work needed to produce these figures might well have lengthened into the work of many weeks.

Design Evaluation

In the course of the design exercise, a number of interesting properties of large quad arrays emerged. Some of the patterns making up the properties might not have been so easily discovered without the efficiency of computer-aided design, although an automated design procedure might have obscured some of them. Let's analyze the designs band-by-band. We shall use a mixture of tabular and graphic data to examine each band.

20 Meters

All of the large arrays show a steady increase in gain with each step upward in frequency band. Of all the bands, 20 meters shows the least improvement as we enlarge the array. Table 2 provides the data for the band edges and in the middle of the band. Fig 6 sweeps the band to provide a complete picture of the free-space gain.

At the low end of the band, the 4-5 array shows a significant increase over the 3-4 version. The gain increase tapers off as we move up the band so that the average gain margin between versions 3-4 and 4-5 is the same as between version 4-5 and 4-5-6. However, the gain of ON7NQ's version 3-4 had been optimized at the expense of full-band coverage. Both 4-5 and 4-5-6 provide full coverage of 20 meters, even if at lesser levels at the high end of the band.

The front-to-back ratio curves for all three quad versions appear in Fig 7. The curves are roughly congruent, but the increasing boom length of the array as we move from one version to the next yields a higher peak value at about 14.1 MHz. Although both larger arrays have a higher ratio than the original ON7NQ array at the low end of the band, all three pass the upper end of the band with similar values. Likewise, as shown in Fig 8, the two larger arrays have similar SWR curves that barely fit within the band at less than 2:1 SWR relative to 50 Ω. In this feature, they are superior to the original ON7NQ array, since its SWR curve cannot be moved sufficiently to cover the entire band without a significant reduction in peak gain.

Adding a second director to the initial ON7NQ thus allows an improvement of gain of modest amounts. The added director permits coverage of the entire 20-meter band by judicious selection of director loop sizes, which differ as we change the boom length. Without major changes in individual element spacing, further performance improvement is unlikely, since the first director has two functions. In combination with the driver and the reflector, it sets the feed-point impedance. In combination with the second director, the first director sets the operating

Table 2—20-Meter Performance of Three Quad Designs

Frequency (MHz)	*Gain (dBi)*	*Front/Back (dB)*	*Impedance (R ± jX)*	*50-Ω SWR*
ON7NQ 3-4-Element, 5-Band Quad				
14.0	8.42	11.83	37.6 –*j*18.5	1.66
14.175	8.29	15.06	44.3 + *j*4.4	1.17
14.35	8.06	9.76	34.8 + *j*36.5	2.50
W4RNL 4-5-Element, 5-Band Quad				
14.0	8.81	15.02	33.6 –*j*20.5	1.88
14.175	8.58	16.76	51.9 + *j*10.0	1.22
14.35	8.14	9.96	57.8 + *j*33.8	1.89
Average gain over 3-4: 0.25 dB				
W4RNL 4-5-6-Element, 5-Band Quad				
14.0	9.04	15.37	31.7 –*j*18.4	1.89
14.175	8.82	17.82	54.9 + *j*12.8	1.29
14.35	8.41	10.36	56.6 + *j*35.3	1.94
Average gain over 4-5: 0.25 dB. Average gain over 3-4: 0.50 dB				

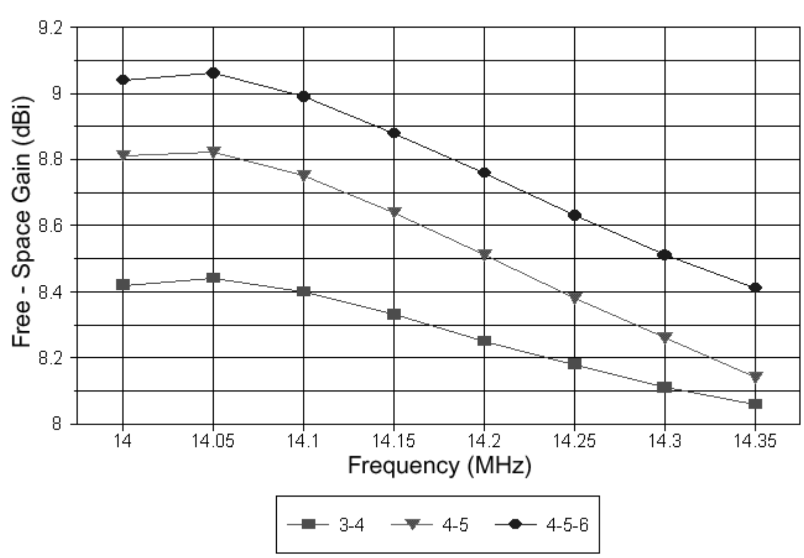


Fig 6—20-meter free-space gain for three large five-band quad designs.

bandwidth for the major parameters. Hence, the two four-element 20-meter designs use very different director sizes, although the driver and reflector remain constant. All in all, both larger 20-meter sections have boom lengths that remain well under 0.4 λ, which is short for a four-element parasitic array.

17 Meters

Because 17 meters is such a narrow band (100 kHz), the data in Table 3 will suffice to permit an evaluation of the performance of the arrays on this band. Three factors allow the 17-meter portions of the larger arrays to achieve significant gain over the initial three-element quad. First, the boom length increases as a fraction of a wavelength so that the two new sections bracket a half wavelength of boom length. Second, the extra element is well suited to setting the element mutual coupling for a higher gain level. Third, the 17-meter band is narrow, permitting the operating performance to be well focused.

Nevertheless, the 30-foot-boom version requires different lengths than the 26-foot version for the two directors to achieve the final 0.2-dB gain increment. However, the added length also permits the designer to obtain feed-point impedances closer to 50 Ω, even though both four-element designs have comparable F/B values.

Despite the factors that allow the 17-meter section to achieve gain in excess of the specifications for the larger arrays, the gain differential between the 17-meter and 20-meter sections calls for brief comment. The longer boom length (in terms of a fraction of a wavelength) and the narrow band requirements on 17 meters contribute to the gain excess over that at 20 meters. The 17-meter elements in their planar supports are surrounded on both sides by elements for other bands. Changes in the 20-meter and 15-meter elements do affect the performance curves of 17 meters—much more of an effect than changes to the 17-meter elements have upon the 20-meter performance curves. In general, being surrounded by elements for other bands tends to improve gain, but this also tends to slightly reduce the F/B and SWR bandwidth.

15 Meters

As shown in Table 4 and in Fig 9, 15 meters is marked by remarkable gain stability for all three quad versions. The gain improvement for the 4-5 array over the 3-4 array is more than 1 dB, with another ⅓ dB added by the increased boom length of the 4-5-6 quad. However, these values, which are in excess of expectations for the 4-5-6-element design, required the addition of a new director six feet between the first and second directors for 20 and 17 meters. Table 4 shows the best gain values obtained with the longer boom, with and without the added director. Obviously, the longer boom—about ⅝ λ—was insufficient to provide stable gain across the band without an intervening director.

The front-to-back curves for 15 meters, shown in Fig 10, are equally interesting. The initial 3-4 array, with a single director for 15 meters, shows the typical "spike" in the maximum front-to-back value. Both longer boom models provide much smoother performance across the entire band. The smoother performance is also reflected in the feed-point impedance values. The 3-4 array can be adjusted for less than 2:1 SWR across the band, but it cannot approach the leveled values for the longer-boom arrays.

Part of the reason for the impedance and SWR situation is revealed in Fig 11, the 50-Ω SWR curves for the three arrays. The 3-4 array shows the typical curve of a three-element beam, with a single SWR minimum. Both the 4-5 and the 4-5-6-element arrays display two SWR minimums at different points within the band. The double-dip

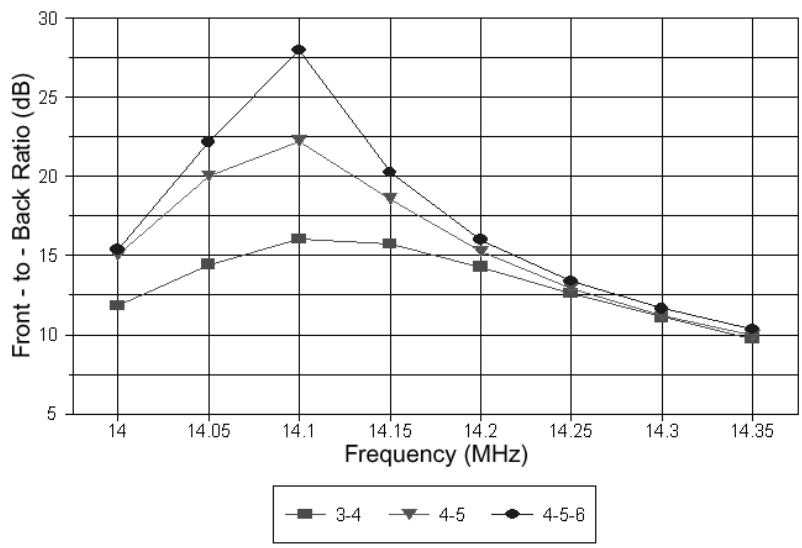


Fig 7—20-meter front-to-back ratios for three large five-band quad designs.

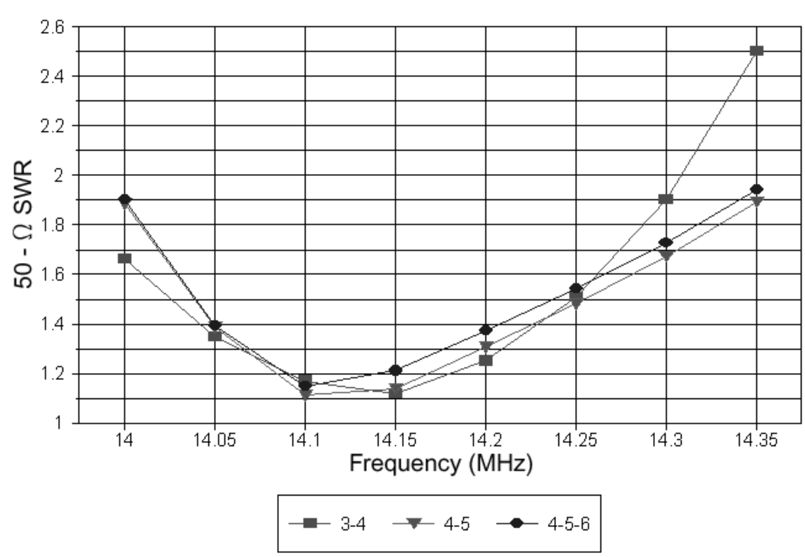


Fig 8—20-meter 50-Ω SWR curves for three large five-band quad designs.

curve is a mark of "wide-band" tuning of a parasitic array, as the term is defined in the series of Yagis designed by NW3Z and WA3FET.[4] The 15-meter array spaces the reflector and the first director at nearly optimal distances from the driver to set the feedpoint impedance for wide-band 50-Ω operation. The reflector is about 0.216λ behind the driver, while the first director is about 0.173λ ahead. An additional director or directors then provide gain, as shown in Table 1. The additional director in the 30-foot model permits the designer to achieve smoother wide-band performance in all categories at a boom length that is beyond the capabilities of a single director.

Both larger quads call for a smooth decrease in the sizes of the directors, although some wide-band applications with five elements may require that the second director length be equal to or slightly larger than the first director. This phenomenon is an indication that the forward two directors are the principal determinants of the bandwidth for the gain and F/B curves.

12 Meters

For all three arrays, 12 meters is the lowest band to use a driver spaced five feet from the reflector, with the element at the 10-foot mark becoming a director. Table 5 provides the operating figures for this five-element section. The 4-5 (26 foot) array provides over 1 dB gain improvement over the 3-4 (18 foot) array on 12 meters, with similar F/B and SWR curves.

When increasing the boom length to 30 feet, the forward director moves from 0.2 to 0.3 λ ahead of the second director. The larger spacing is near the limit of the ability of the forward two directors to control both gain and front-to-back ratio, even on a narrow (100 kHz) band such as 12 meters.

Indeed, without a further director, one can improve either the gain or the smoothness of F/B—but not both. As shown in Table 5, the design approach used for the 4-5-6 array was to raise the lowest level of F/B by about 2 dB—a value just verging on operational detectability.

The gain of the longer-boom array increases insignificantly over that of the 26-foot model. However, since the overall gain increase was in excess of 1 dB relative to the initial 3-4-element array, this design decision seems appropriate. For the same reason, I didn't add a new director to the support arms used for the added 15-meter director. The absence of an added director for 12 meters illustrates once more the different requirements for narrow and wide amateur HF bands.

10 Meters

Of all the HF bands, 10 meters is the widest. From the outset, it was apparent that a thin-wire quad array could not cover even the full first megahertz of 10 meters adequately. An 800-kHz operating bandwidth is a much more feasible goal, and it is achieved by all three arrays, as shown in Table 6.

On average, the 4-5-element quad shows better than 0.8 dB more gain than the 3-4 array. The 30-foot boom model adds more than 0.4 dB more gain (using a fourth director), for a 1.25-dB total improvement over the original 18-foot quad design. However, these figures—as averages—may be deceptively simple in view of the wide operating bandwidth on 10 meters.

Despite the best efforts to achieve a smooth gain performance, 10 meters exhibits the highest differential between minimum and maximum gain for all three of the arrays. The differential runs between 0.9 dB and 1.1 dB,

Table 3—17-Meter Performance

Frequency (MHz)	*Gain (dBi)*	*Front/Back (dB)*	*Impedance (R ± jX)*	*50-Ω SWR*
ON7NQ 3-4-Element, 5-Band Quad				
18.068	8.47	21.80	42.7 –*j*5.1	1.21
18.118	8.42	25.52	43.5 –*j*0.3	1.15
18.168	8.36	20.90	43.2 + *j*4.6	1.19
W4RNL 4-5-Element, 5-Band Quad				
18.068	9.24	22.03	36.0 –*j*1.7	1.39
18.118	9.18	21.26	39.3 + *j*5.7	1.31
18.168	9.10	17.39	42.3 + *j*12.5	1.37
Average gain over 3-4: 0.75 dB.				
W4RNL 4-5-6-Element, 5-Band Quad				
18.068	9.45	18.43	42.7 + *j*0.7	1.17
18.118	9.39	21.33	47.9 + *j*6.5	1.15
18.168	9.31	20.50	42.4 + *j*10.8	1.24
Average gain over 4-5: 0.21 dB. Average gain over 3-4: 0.96 dB				

Table 4—15-Meter Performance of Three Quads

Frequency (MHz)	*Gain (dBi)*	*Front/Back (dB)*	*Impedance (R ± jX)*	*50-Ω SWR*
ON7NQ 3-4-Element, 5-Band Quad				
21.0	8.43	15.28	49.7 –*j*20.1	1.49
21.225	8.52	20.98	46.4 –*j*0.0	1.08
21.45	8.47	10.24	36.2 + *j*30.7	2.16
W4RNL 4-5-Element, 5-Band Quad				
21.0	9.49	15.33	41.4 –*j*15.6	1.47
21.225	9.47	17.04	57.0 + *j*7.5	1.21
21.45	9.55	19.16	31.3 + *j*9.9	1.70
Average gain over 3-4: 1.03 dB.				
W4RNL 4-5-6-Element, 5-Band Quad (before adding fifth element)				
21.0	9.36	11.43	46.4 –*j*19.8	1.51
21.225	9.65	22.65	58.1 –*j*8.8	1.25
21.45	9.95	15.68	28.2 + *j*8.7	1.85
W4RNL 4-5-6-element, 5-Band Quad (after adding fifth element)				
21.0	9.78	15.70	46.9 –*j*7.6	1.19
21.225	9.74	20.57	63.4 + *j*1.0	1.27
21.45	10.00	15.03	35.0 + *j*11.9	1.57
Average gain over 4-5: 0.34 dB. Average gain over 3-4: 1.37 dB.				

depending on the version of the array. The 4-5-6-element array would have shown an unacceptably high differential—more than 1.5 dB—had the final design not included a new director on the same support arms as the added 15-meter director. Fig 12 shows the gain curves for all three final designs. Even with the new director, the 4-5-6 version displays a more rapid gain fall-off at the upper end of the band than the other two quads.

The F/B curves in Fig 13 show something about where to place the peak F/B value for optimal performance—if there is design room to vary it without adversely affecting other properties. The 3-4-element array centers the curve. The 4-5-element version moves the maximum 180° F/B to the upper end of the band. The result is less performance at the lower end of the band. However, the 180° F/B is not the sole determinant of placement. The general shape of the rearward lobes and the strength of rearward side lobes can also play a role in the design decision. Placing the maximum F/B ratio at the high end of the band in the 4-5 array provided the best F/B performance across the band.

The addition of another director to make the 10-meter section a six-element array was prompted by the F/B performance as much as by the gain curve of the array. Table 6 shows the high in-band, peak F/B without the new director. The consequence is relatively poor F/B performance except for a small portion of the band. With the added director, the F/B performance curve spreads the higher levels of performance over a greater portion of the band, although the band edges fall below the target levels in the specifications.

The added 10-meter director also resolves another problem. Without the director, the elevated SWR between the two wide-band minimums rises too high and exceeds the 2:1 level by a considerable amount at 28.6 MHz. As shown in Fig 14, all three final versions of the arrays achieve the double-minimum wide-band curve, although in different patterns. The 4-5-6-element 30-foot array achieves the flattest curve of all, but all three curves remain below the 2:1 SWR level for the entire operating bandwidth. In both the 4-5 and 4-5-6 arrays, the first director plays its most significant role in setting the feedpoint impedance of the array and hence turns out to be smaller than the second director.

Overall Evaluation: The 4-5-element and the 4-5-6-element 5-band quads achieve most of the operating goals set forth in the original specifications for array gain. Each array exhibits increased gain as we change bands upward in frequency. Only 12 meters fails to provide at least 0.2 dB more gain for the 4-5-6 array over the 4-5 version. Only 20 meters fails to meet the goal of the 4-5 quad in providing at least 0.7 dB more gain than the array used as the starting point.

The F/B goals—with cautions that we shall further discuss—are generally met, except at the upper end of 20 meters and the passband edges of 10 meters. Both the 4-5-element and the 4-5-6- element quads cover all of the assigned passbands with less than a 2:1 SWR relative to a 50-Ω standard. However, in several cases the limit is pressed on one or the other end of the band, and on both ends of the 20-meter band.

Within these restrictions, then, the design is reasonably successful in achieving a design for a larger five-band quad array. Indeed, more important than this evaluation are the design principles and limitations discovered along the design road. The notes on these matters give us further insight into how multielement, multiband quads operate.

Moreover, it is critical to understand that the designs emerged from some initial constraints of wire size and fixed element spacing. Revising the element spacing among sets of elements might

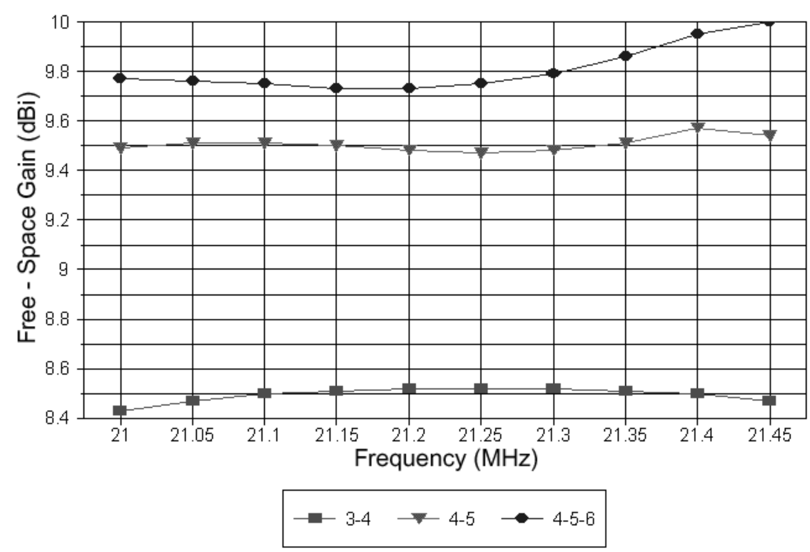


Fig 9—15-meter free-space gain for three large five-band quad designs.

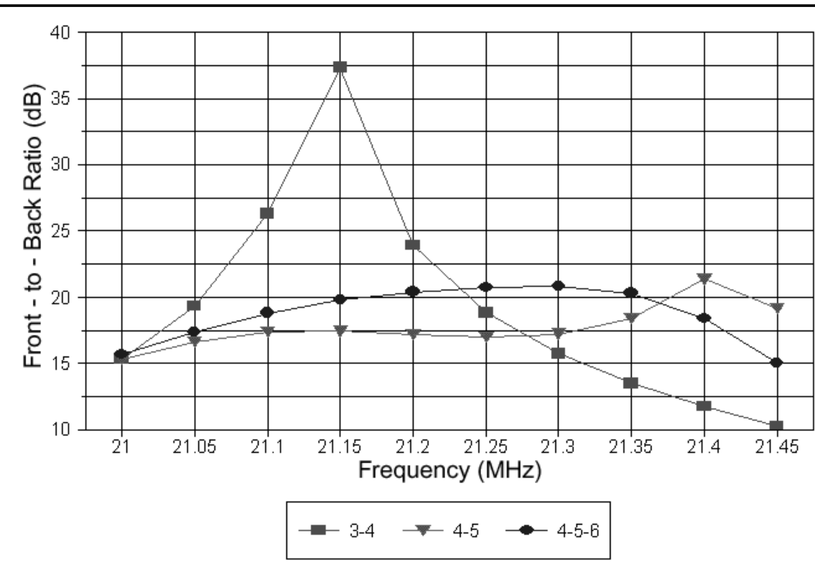


Fig 10—15-meter front-to-back ratios for three large five-band quad designs.

yield differences in F/B curves and SWR curves. Consequently, the designs are now suited to the application of both incremental and genetic optimizing routines. The incremental routines may provide further tweaking of the element sizes in the direction of perfected performance curves. Genetic algorithms might well uncover unsuspected potentials for the array designs.

Even if we accept the declaration of relative success in designing 4-5-element and 4-5-6-element quad arrays, the design process is far from over. First, there are some further general cautions about multiband quads that deserve to be addressed. Second, although the use of antenna modeling software shows good efficiency in developing a design, if the design cannot be translated into an effective physical antenna with adequate performance, the exercise is somewhat futile!

Limitations, Cautions and Correlation Techniques

The numbers that emerge from an antenna-modeling program used to design a large quad array can be deceptive, unless we use extreme caution in reading them and in using them to construct a physical version of the array. In this final part of the exercise, I want to explore at least some of the limitations and cautions that attach to the design model and its transferal into wire and fiberglass.

Patterns

The basic design work has been done with free-space models. Hence, all gain figures require readjustment relative to a proposed height for the array above a specific ground quality. Ordinarily, the F/B or rearward lobes and the SWR curve will hold if the array is more than $\lambda/4$ above ground. Quads are less sensitive to ground influences on the feed-point impedance and other operating characteristics than are arrays with open-ended linear elements. The exact gain of the strongest lobe and the elevation angle of that lobe will, of course, be functions of antenna height, as measured in wavelengths above ground.

With the exception of quite low mounting heights, azimuth patterns over real ground will closely resemble at the elevation angle of maximum radiation the free-space patterns from the design model. However, we might need to adjust our expectations for such patterns due to the high levels of interaction among the elements of a multiband quad. Not all bands produce the clean patterns we have come to expect from monoband Yagis.

Fig 15, for example, might represent both the 20-meter and 17-meter bands for the 4-5-6-element array. In both cases, we have well-behaved patterns, with single forward lobes and no forward side lobes. We also have radiation to the rear that follows fairly standard progressions: showing three small rearward lobes, a single broad lobe or something in-between the two. However, even the pattern for 18.168 MHz reveals a good reason for the quad designer to look at each pattern over several frequencies. The rearward pattern at the upper end of 17 meters shows a worst-case front-to-rear ratio of about 17.5 dB, despite a 180° F/B of better than 20 dB.

Above 17 meters, the patterns—both forward and rearward—can grow considerably less well behaved. In the progression from the middle of 15 meters to the upper end in Fig 16, we find the development of forward side lobes. Although they remain diminutive at 15 meters, on higher bands, the side lobes can grow to proportions that affect the overall forward beamwidth of the array between –3 dB power points. In addition, the large rearward radiation pattern, with a worst-case ratio to the forward lobe of 15 dB may have operational consequences. This is because the response to the rear would no longer be in a pair of narrow directions, but instead would cover most of the rear quadrants. In your preconstruction evaluation of a large quad design, you

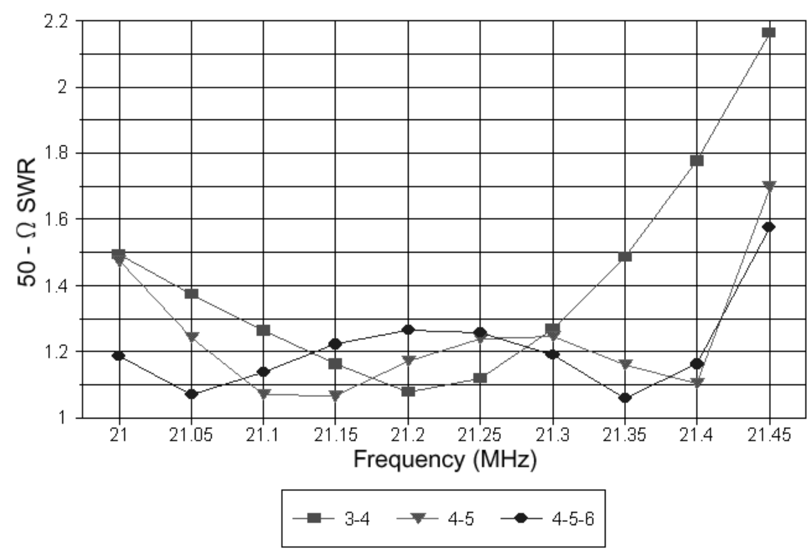


Fig 11—15-meter 50-Ω SWR curves for three large five-band quad designs.

Table 5—12-Meter Performance of Three Quads

Frequency (MHz)	*Gain (dBi)*	*Front/Back (dB)*	*Impedance (R ± jX)*	*50-Ω SWR*
ON7NQ 3-4-Element, 5-Band Quad				
24.89	9.26	22.72	35.1 –*j*2.1	1.43
24.94	9.22	18.92	41.1 + *j*2.3	1.27
24.99	9.18	16.70	47.6 + *j*4.8	1.12
W4RNL 4-5-Element, 5-Band Quad				
24.89	10.27	21.77	38.6 + *j*5.2	1.33
24.94	10.29	19.80	40.2 + *j*9.1	1.34
24.99	10.25	16.77	41.9 + *j*14.3	1.43
Average gain over 3-4: 1.04 dB.				
W4RNL 4-5-6-Element, 5-Band Quad				
24.89	10.34	18.78	25.9 + *j*3.5	1.94
24.94	10.37	20.98	37.0 + *j*7.9	1.42
24.99	10.25	21.69	49.1 –*j*2.5	1.06
Average gain over 4-5: 0.06 dB. Average gain over 3-4: 1.10 dB				

should decide whether or not the patterns (as well as the performance numbers) are satisfactory for your intended operation.

Efficiency

The *NEC* core at the heart of most antenna-modeling software packages provides a power budget that lists a value for efficiency. The efficiency of an antenna is simply the power radiated (without regard to where it goes) to the power supplied to the antenna, expressed as a percentage. The calculation does not include anything not modeled; for example, matching sections or networks, feed-line losses, and so on. However, it does include material losses within the antenna elements as a result of their resistivity, and it also includes resistive losses associated with any traps or reactive loads. This latter category of losses does not apply to our quad arrays, but wire losses do apply, since we are using #12 AWG copper wire. The wire size is as important as the material, since skin effect is partially a function of element surface area. In fact, with large element surface areas, such as with the use of aluminum tubing, material losses can be very small. For example, I have models of six-element Yagis in my collection with efficiencies approaching 99%.

Thinner wire (as a fraction of a wavelength) and higher frequencies increase losses and lower the efficiency of an antenna. These general rules would reveal themselves if we developed a sequence of simple monoband Yagis by which to test them. However, the large quads we have been exploring display complex interactions among the elements. In doing so, they reveal another dimension to antenna efficiency that is not as well appreciated as element diameter and frequency.

Table 7 lists the calculated mid-band efficiencies of each of the quads reviewed. Notice that the highest efficiency is considerably lower than that for a “fat-element” Yagi. Although we can detect a pattern in the general direction of changes in efficiency, there are some surprises. Especially noticeable is the very low efficiency figure for 12 meters for the largest array.

If we return to Table 5, we discover that the 12-meter portion of the largest array provided less than 0.1 dB gain advantage over the next shorter quad, with most other characteristics being roughly equal between the two. What limits gain is the inability of the elements on fixed spacings to achieve the most effective interelement coupling to yield a higher gain. If the larger array had resulted in significantly larger rear lobes, the efficiency might actually have been higher. Had it resulted in higher forward gain—or even a wide beamwidth—we might also see a higher efficiency value. However, we often neglect a third possibility: the current distribution in all elements is such that the sum of radiation in all directions does not increase, but instead, the current levels are higher in regions of the antenna where losses exceed contributions to radiation. The result is lower efficiency without a change in wire size, wire lengths or frequency.

For the 12-meter case, we might raise efficiency somewhat by adding a fourth director (as was done on 10 meters), even though it would add to wire losses. We might optimize further the relative spacing of the 12- and 10-meter drivers from the reflector.

Efficiency is (or can be) an indicator of possible design improvement. However, it does not affect the reported gain of the array, since that gain already takes into account the radiation efficiency of the total antenna model. Indeed, attaching the wrong significance to efficiency can result in a misuse of the data. For example, achieving 99% efficiency in a directional array, where the added radiation is to the rear or sides, would not amount to a design improvement.

Element Precision

An array with highly interactive parasitic elements requires consider-

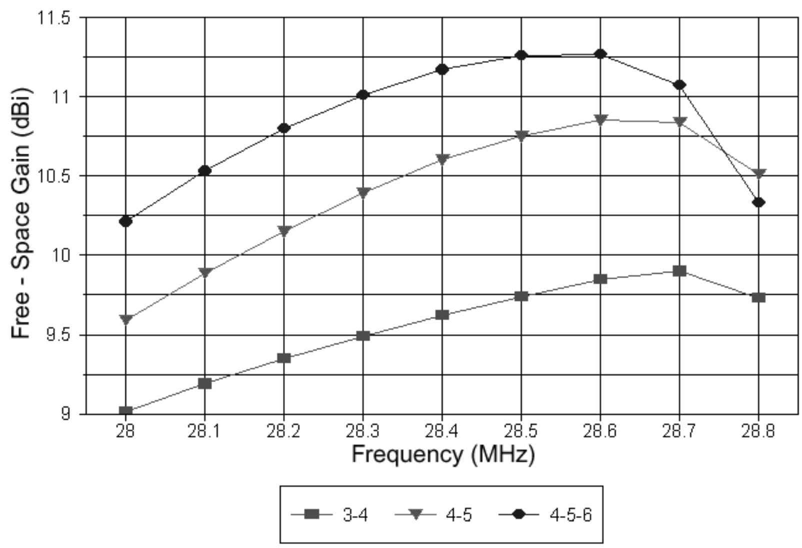


Fig 12—10-meter free-space gain for three large five-band quad designs.

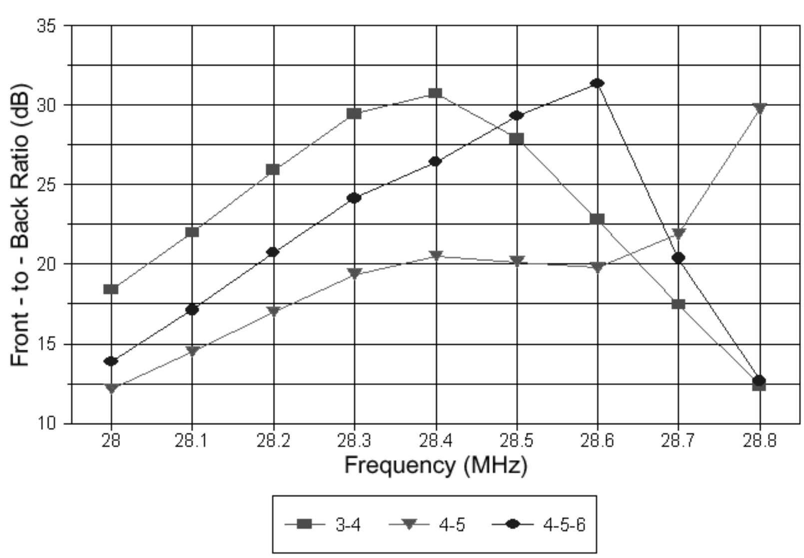


Fig 13—10-meter front-to-back ratios for three large five-band quad designs.

able precision in construction to achieve the design results. One aspect of construction precision is understanding which elements can be adjusted and which should be precisely built and then left alone. The following guidelines may be useful, although their application may vary from one design to another.

Reflector-driver-first director: For arrays with two or more directors on a band, fine tuning of the reflector-driver-first-director combination tends to set the source impedance across the band in question. Once set, you should not perform further adjustments on this set of elements—with one exception. The driver loop can be adjusted to tune out reactance at the feed point. However, reductions in driver size will normally also reduce the resistive component of the feed-point impedance, and increases in size will raise the feed-point's resistive component. In constructing a given large array, adjustments here should be done last.

First director: Where there are two or more directors, the size of the first director may be sufficiently critical that, once set you should not alter it. On some bands, less than a 1-inch change in the first director can create large changes in the performance within the passband, involving any of the key operating parameters: gain, F/B or SWR curve. In general, the further forward along the boom you make element changes, the less critical they are.

The two forward-most directors: As Table 1 reveals by comparing dimensions among arrays, you can go far toward controlling the characteristics on a band by changing the forward directors. For wide-band service, the most-forward director becomes shorter to enhance high-end performance and the next director to the rear becomes larger to enhance low-end performance. Both moves tend to raise the feed-point resistive component a bit, which is why hasty adjustments to the driven element should be avoided.

Obviously, where there are too few elements to adhere to these guidelines, you will need to employ other measures. For example, a band with four elements may require a slight enlargement of the reflector to enhance low-end performance. However, this move may require re-adjustment of the driver and first director to restore or obtain desired feed-point impedance and the SWR curve across the whole band. A three-element band becomes a real ballet of interactions among the elements, such that the reflector is normally used to control both radiation resistance and low-end performance, while the director controls high-end performance, with the driver sized to create the best possible situation for the antenna feed. Since large multiband quad arrays normally begin with fixed spacing, there are many instances where meeting all design specifications may not be possible.

Table 6—10-Meter Performance of Three Quads

Frequency (MHz)	*Gain (dBi)*	*Front/Back (dB)*	*Impedance (R ± jX)*	*50-Ω SWR*
ON7NQ 3-4-Element, 5-Band Quad				
28.0	9.01	18.40	43.8 –*j*31.6	1.96
28.2	9.35	25.89	45.3 –*j*11.0	1.29
28.4	9.62	30.72	51.3 + *j*6.8	1.15
28.6	9.85	22.80	58.7 + *j*9.6	1.27
28.8	9.73	12.38	31.1 + *j*8.1	1.68
W4RNL 4-5-Element, 5-Band Quad				
28.0	9.59	12.15	40.7 –*j*27.4	1.88
28.2	10.15	17.00	49.3 –*j*12.7	1.29
28.4	10.60	20.50	47.1 –*j*2.8	1.09
28.6	10.85	19.76	42.6 + *j*18.0	1.52
28.8	10.51	29.74	64.9 + *j*12.1	1.40
Average gain over 3-4: 0.83 dB				
W4RNL 4-5-6-Element, 5-Band Quad (before adding sixth element)				
28.0	9.54	18.64	41.3 –*j*19.4	1.59
28.2	10.22	43.18	40.0 –*j*1.3	1.25
28.4	10.72	20.43	39.2 + *j*23.6	1.78
28.6	11.04	16.89	59.1 + *j*55.3	2.70
28.8	10.67	11.37	56.4 –*j*19.1	1.46
W4RNL 4-5-6-Element, 5-Band Quad (after adding sixth element)				
28.0	10.21	13.85	58.9 –*j*31.2	1.80
28.2	10.80	20.75	51.9 –*j*21.9	1.54
28.4	11.17	26.41	47.1 + *j*0.6	1.06
28.6	11.27	31.30	62.3 + *j*16.0	1.43
28.8	10.33	12.67	34.0 + *j*1.7	1.47
Average gain over 4-5: 0.42 dB. Average gain over 3-4: 1.25 dB				

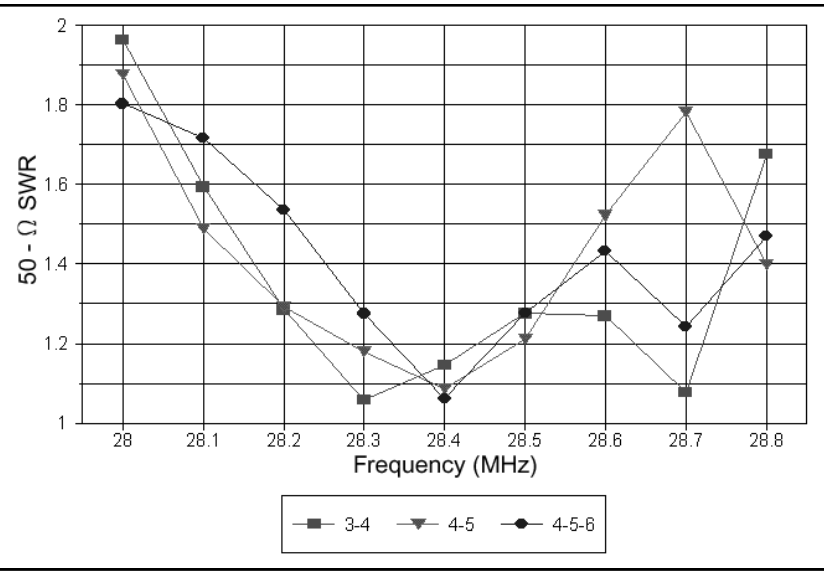


Fig 14—10-meter 50-Ω SWR curves for three large five-band quad designs.

Computer Models and Reality

Because many dimensions within a large multiband quad array require precise measurement of the loop circumference to within less than one inch, constructing such an array is not a casual endeavor. Indeed, it may lie beyond the realm of simple backyard build-and-play techniques. However, with some care, trials and testing, constructing a modeled design is still feasible. The key lies in understanding both the model and the realities of a proposed construction technique.

Fig 17 sketches loosely some of the ways builders attach quad-loop corners to the support arms. In two of these, we see metal rings wrapped around the nonconductive support arm. In some cases, the element wire may be wrapped at the corner to reduce abrasion. Whether or not directly connected, we can end up with a small one-turn coil in close proximity to the quad-loop corner. The current at a quad loop corner is significant, in fact higher than the current magnitude on a linear element the same distance from center. The closed loop may function as a load on the quad loop, and it

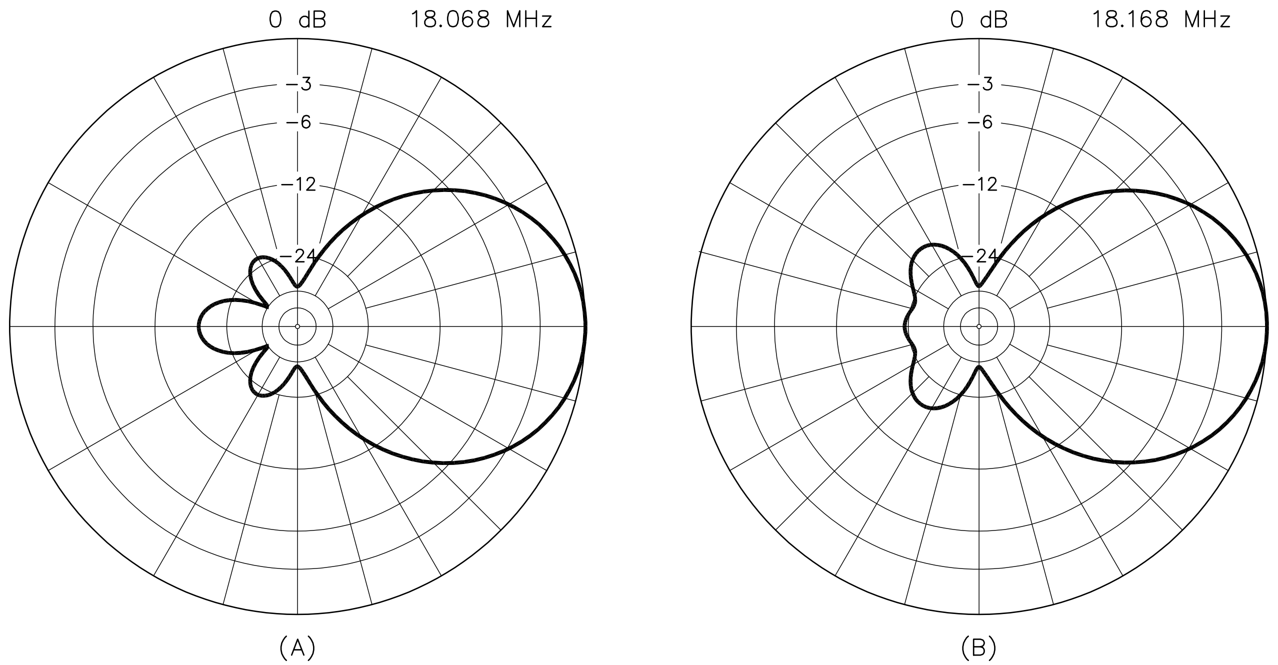


Fig 15—Sample well-behaved free-space azimuth patterns from 17 meters.

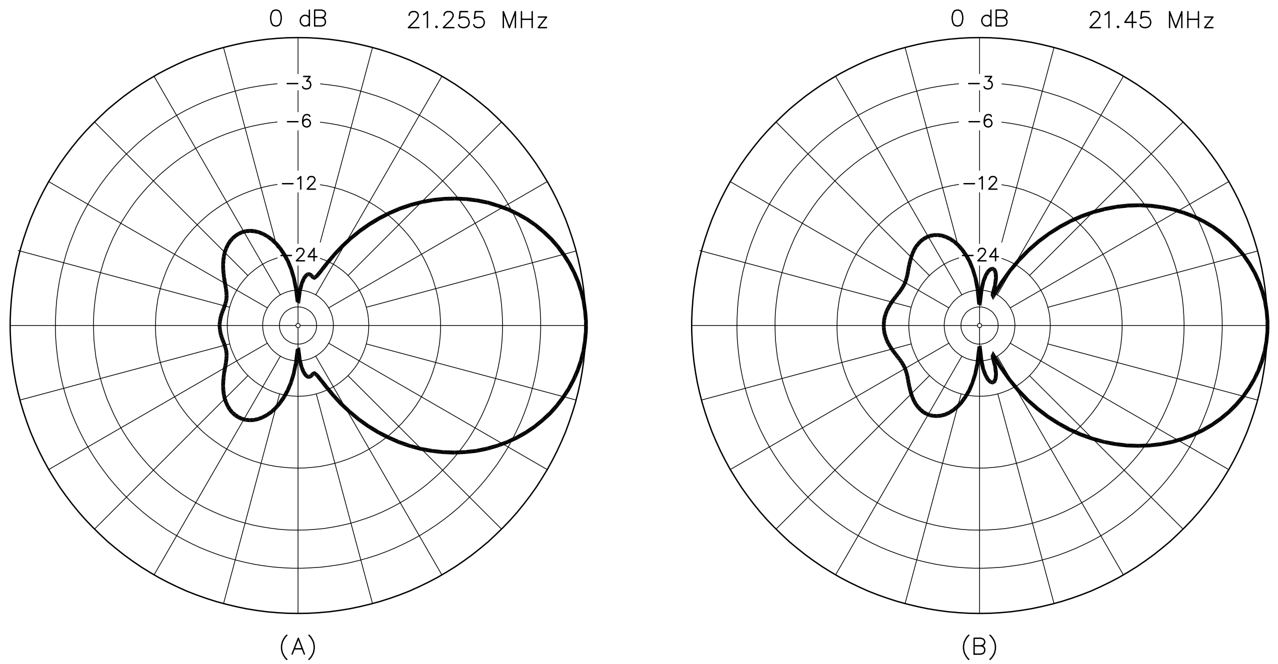


Fig 16—Sample less well-behaved free-space azimuth patterns from 15 meters.

does not take much of a load to detune the element relative to the original model of the loop.

Similarly, when quad arms are composed of combinations of aluminum and nonconductive material, the aluminum may be close enough to the loop corner to create a slight detuning. This is also equivalent to adding a very small load to the wire loop. Of the four methods for wire attachment to a support arm, the nonconductive cable-tie system comes closest to matching the computer model.

Some users employ an anti-abrasion sleeve over the element wire, making in effect a short piece of insulated wire. The insulation causes a velocity factor, making the physical and electrical length of that portion of the wire unequal. Corner sleeves may turn out to be harmless relative to the complex operation of a large multiband array, but they should not be presumed to be harmless.

There is a tedious, but straightforward, way you can determine the degree to which construction practices affect the operation of a quad relative to the "clean" bare-wire computer model on which it is based. First, model only the driven element assembly or assemblies. Determine as precisely as feasible the resonant frequency for each driver. Second, build as precisely as possible the driver assemblies using your preferred method of construction and elevate them to a good height. Now determine the actual resonant frequency for each driver. Either you'll be lucky, and resonant frequencies will match those of the model, or a pattern of offset will become evident. If you find offsets but no pattern, this will likely be good reason to review your initial driver construction.

Relative to the measured resonant frequencies, add identical reactive loads to each of the four corners of each driver so that the model resonates at the same frequency as the driver assembly tested. For each band, adding the same loads to each of the four corners of quad loops on that band will be an accurate representation of the effects your construction techniques have on the entire set of elements. Now, with the added loads, readjust the dimensions of the model to restore the performance curves of the original model. The resulting dimensions should result in correct operation of the array on all frequencies.

I can only state that they "should result in correct operation," but the effectiveness of this technique will rest upon the precision with which you construct your array during both the test and final construction phases of the operation. Even small amounts of loading or detuning on some elements may throw the array off the desired performance curve on some bands.

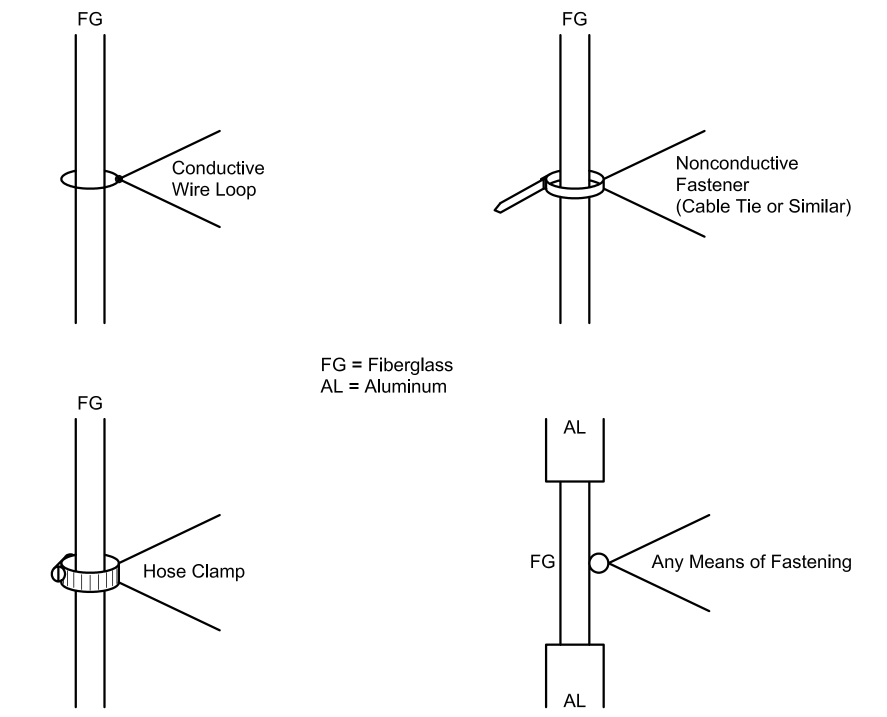


Fig 17—Sample element-to support mounting techniques.

Table 7—Radiation Efficiencies of Three Large Quad Arrays

Band (m)	*Frequency (MHz)*	***Antenna Efficiency (%)*** *3-4-Element*	*4-5-Element*	*4-5-6-Element*
20	14.175	93.6	94.6	94.4
17	18.118	93.9	92.7	93.3
15	21.225	94.1	92.7	93.7
12	24.94	90.1	87.7	80.4
10	28.4	93.6	91.9	90.7

Note: Efficiency is the ratio of power radiated to the power supplied to the antenna and does not include matching or line losses.

Variations of the technique suggested here for correlating modeled quads and physical quads are adaptable to many other types of antennas. More important is the general thesis that models—usually using bare wire and with no modeled detuning effects—require correlation to the physical construction methods employed by the builder if the models are to be adequate guides to antenna design. Any success in building a large multi-element, multiband quad of the order discussed in these notes will depend upon this step as much as any other in the design process.

The design of a large multiband quad array intended for eventual construction can be enhanced by the proper use of antenna modeling software. However, as we have seen, the task is not a mere modeling exercise. It must be preceded by careful consideration of constraints, specifications and modeling strategies to ensure reasonable results. Moreover, the task is not complete unless the final design model is carefully evaluated and then correlated to the proposed construction methods. These notes have had as their goal to make the process orderly, but by no means brief.

Notes

[1]*Quad Notes*, Vol 2 (Corpus Christi: AntenneX, 2001), throughout.

[2]*Quad Notes*, Vol 1 (Corpus Christi: AntenneX, 2000), Chapter 5.

[3]*Quad Notes*, Vol 1 (Corpus Christi: AntenneX, 2000), pp 206-216. See also Danny Mees, ON7NQ, "Improving the Cubex Three-Element, Five-Band Quad," *The ARRL Antenna Compendium*, Vol 6, pp 119-20.

[4]For a description of the set of "optimized wide-band antennas" or OWA Yagis, explore the following Web site: **nw3z.contesting.com**. □□

An All Digital Fractional-N Synthesizer

Designing a high-performance frequency synthesizer is not a trivial task, especially when modern components such as direct digital synthesizers (DDS) ICs are not available. This synthesizer uses cheap off-the-shelf components, but modern all-digital fractional-N synthesis techniques, that allow it to meet or beat the performance of much more complex and costly designs.

By Oleg Skydan, UR3IQO

During the design of a home-made DSP HF transceiver, I faced the challenge of the first LO synthesizer design. I studied several solutions to improve the performance of synthesizer—to obtain small step size while keeping the other parameters state-of-the-art. A multiple-loop design was too complex and expensive for home building.[1] The hybrid DDS-driven PLL synthesizer discussed in *QST* and *QEX* in recent years[2, 3] would be a suitable solution, but I was unable to get DDS chips. So I decided to try all-digital fractional-N techniques.

[1]Notes appear on page 33.

Litvinova Str 31/8
c. Donetsk-14 Ukraine 83014
skidan@mail.ints.net

After some experiments, I have reached success with a simple all-digital, fractional-N, single-loop synthesizer with fractional spur compensation using a fourth-order sigma-delta modulator.

Fractional-N Divider Basics

Fractional-N synthesizers have been used for many years to improve the performance of indirect frequency synthesizers. Fig 1 shows the principle of a fractional-N divider. The division ratio of the divider is made to have a fractional component by changing the division ratio of the divider periodically, so the average value contains a fractional element. If, for instance, a fractional value of 0.1 is required then the division ratio is changed by one every tenth cycle. If a fractional value of 0.01 is required then the division ratio is changed by one every hundredth cycle. This offers much finer frequency control than integer-based systems.

An accumulator whose digital output is incremented for each cycle of the divider by the fractional frequency requirement is a convenient method of controlling the division ratio. The accumulator uses an adder latch to add the contents of its input to its current output on each cycle of the clock. It behaves as the digital equivalent of an integrator and since the integral of the frequency is phase, its output represents the relative phase of the fractional component. Every time the accumulator reaches its capacity, it produces an overflow, which changes the divider division ratio.

There is a price to be paid for the improvement in frequency resolution these systems provide. The manipulation of the divider ratio generates phase perturbations, and hence, spurious signals that must be eliminated in a useful synthesizer design. The nature of the phase perturbations is predictable and can be cancelled using an analog correction system. Such a system was used in the first LO of the RA1792/RA6790 receivers. Solutions that use various techniques to improve the performance of PLL-based fractional-N systems have been patented.[4, 5] Unfortunately, the fractional-spur suppression was limited by analog circuits.

The Digital Revolution

In 1984, John Wells of Marconi Instruments invented a new solution to the problem of fractional-N synthesizers that did not require analog components to correct noise and spurious emissions normally introduced by fractional-divider schemes.[6] An implementation of such a system is shown in Fig 2.

This digital fractional-N system is based on the principle of noise shaping. Instead of trying to cancel the fractional-N spurs, this system spreads them out over a wide spectrum, and modifies the resulting noise shape so as to minimize low-frequency spectral content (see Fig 3). The technique actively reduces the generation of low-frequency noise and exchanges it for increased levels of higher-frequency noise. This is a very good arrangement because the PLL itself acts as a low-pass filter.

In the example of Fig 2, the single accumulator of the earliest forms of fractional-N synthesizers is replaced with three or more accumulators, the output of each being connected to the input of the next. The overflow from each of the accumulators manipulates the division ratio of the divider.

The first accumulator overflow acts in the same way as the accumulator in the simplest fractional-N systems. It changes the division ratio divider from N to $N + 1$ for one cycle when the accumulator overflows. The remainder output from the first accumulator represents the phase error that would result if no other correction were applied. The second accumulator digitally integrates the output of the first accumulator and subsequent accumulators repeat this process. The overflow from the second accumulator needs to manipulate the division ratio by the differential of the effect of the first accumulator. In a similar way, the output from the third accumulator manipulates the division ratio by the differential of the effect that an overflow output from the second accumulator causes, and so forth.

The sequences that are generated correspond to the terms of a plurality of sequences, each of which represents successive rows in a Pascal triangle (see Fig 4). The sum of each row is zero, with the exception of the first, which is required to correct the division ratio of the divider overall to obtain the required fractional frequency. Because the second and subsequent rows of the Pascal triangles introduce an average divider change of zero, these accumulator overflows have no long-term effect on the division ratio of the divider. However, they are used to remove low-frequency components from the divider's output spectrum and to transfer the energy to higher frequencies where the PLL loop filter can successfully filter them. As shown in Fig 2, the accumulator overflows are fed via delay networks (implemented by D-type flip-flops) to the adder with weighted inputs, so that they generate the required division-ratio changes.

The ability to suppress fractional spurs depends on the initial accumulators' contents and input to the first accumulator. If we want good spur suppression, we should avoid initiation cyclic sequences in the accumulators (like 1000->0000->1000-> and such in a four-bit accumulator), which would shorten the generated sequences.[7] To ensure "longevity," the LSB of the first accumulator input is always set (only odd fractional numbers are used). The effect is the creation of a very long pseudorandom sequence, eliminating initial-condition effects with high-pass characteristics and the desired average

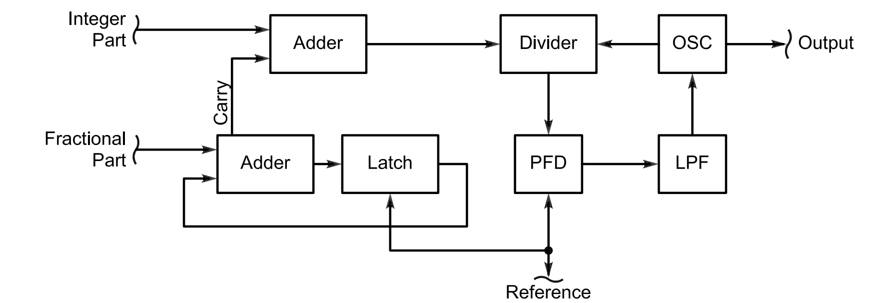


Fig 1—A block diagram of the simple (first order) fractional-N synthesizer.

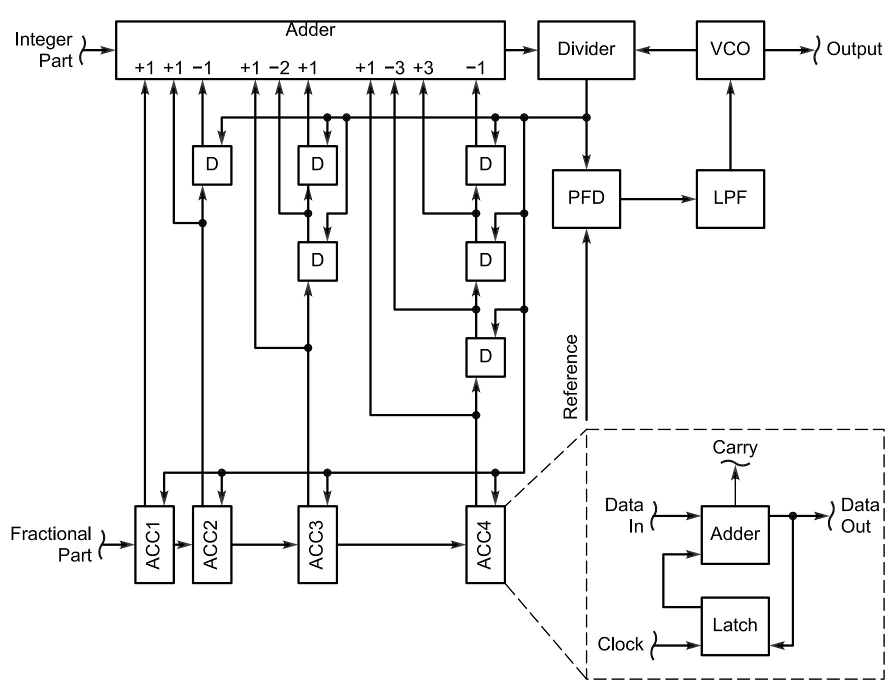


Fig 2—A block diagram of the multiple accumulator fractional-N synthesizer.

division ratio. This can create a frequency error of:

$$\frac{F_r}{2^L} \quad \text{(Eq 1)}$$

where L is the size of the accumulator and F_r is the reference frequency.

Circuit Details

The synthesizer, shown in Figs 5 and 6, consists of the VCO, two isolation amplifiers, an output frequency divider and a PLL circuit. The VCO operates in three bands (see Table 1). Its output is divided by 4 to 14, depending of the band, before feeding it to the first mixer. This improves phase-noise performance.[8] A KP307G junction FET was selected for the oscillator circuit because of its low noise figure. The KP307G is a Russian transistor. A J310 is a suitable substitution for it. A short-circuited line W1 with an unloaded Q of 300 is used as a resonator. The two high-quality PIN-diodes D7 and D8 are used for band switching. (Unfortunately, I do not know a suitable substitution for these Russian parts.) The series-parallel combination of six Varactors is used for better phase-noise performance.

The two amplifiers Q3 and Q4 isolate the output divider and PLL circuit from the VCO. I used a BF998 with a very low reverse-transfer capacitance.

The U7 divides the output of the synthesizer by any value from 1 through 9. Only division ratios from 2 through 7 are used in my transceiver, with an additional divide-by-two stage (74AC74) located on the first mixer board. Thus, the overall division ratio varies among even values from 4 through 14, to obtain square waves for the mixer. A bit unusual: The divider schematic allows avoiding of the additional inverter use.

The heart of the fractional-N synthesizer consists of two ICs (see Figs 5 and 6). The complex algorithm of fractional-N divider and interface with the control processor (via standard SPI bus) are realized by microcontroller U5, while fast logic functions are realized by complex programmable logic device (CPLD) U4. This makes the design simple and low-cost, but the microcontroller's performance (8 MIPS) limits reference frequency to 150 kHz (this

Table 1

VCO Band	*VCO Frequency (MHz)*
1	90.0-98.8
2	85.0-88.0
3	78.3-79.6

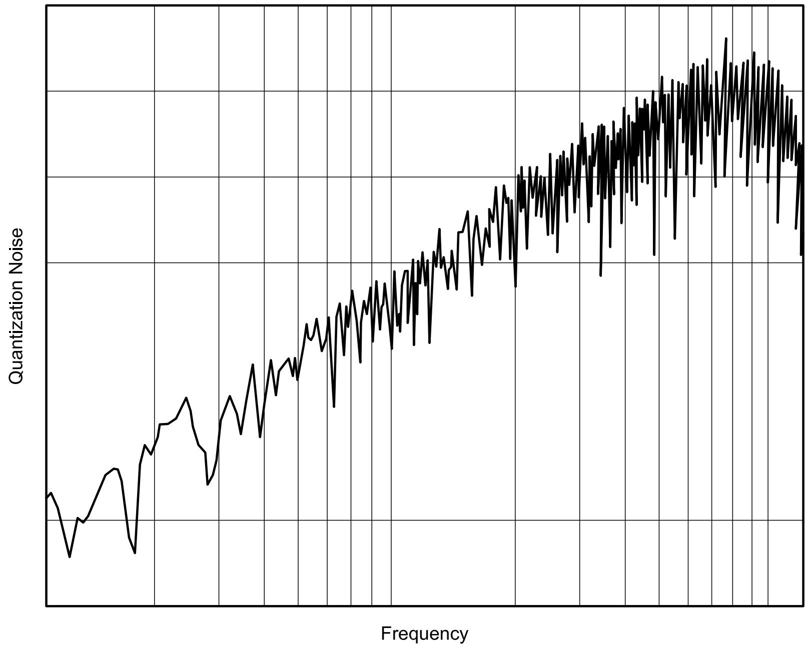


Fig 3—Typical output spectrum shape of the multiple accumulator fractional-N divider.

+1
+1 −1
+1 −2 +1
+1 −3 +3 −1
+1 −4 +6 −4 +1

Fig 4—Pascal's triangle used for division-ratio calculating.

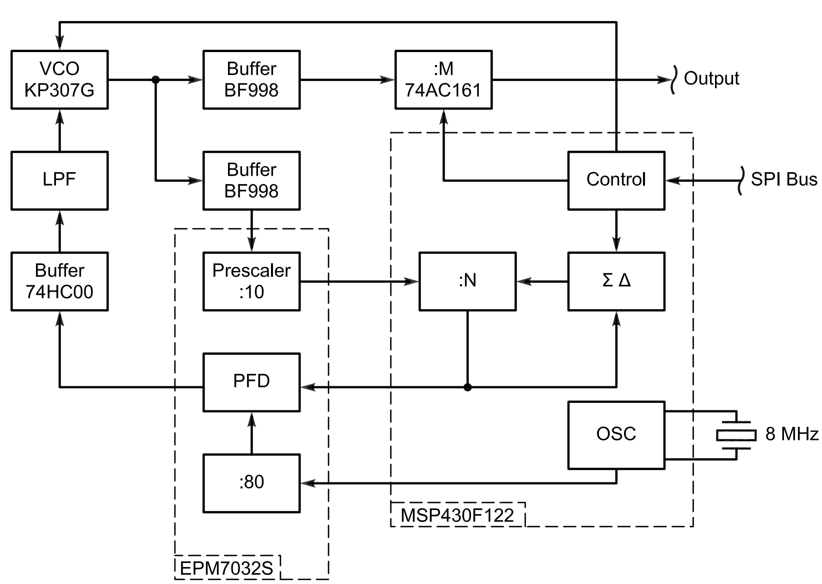


Fig 5—The synthesizer's block diagram.

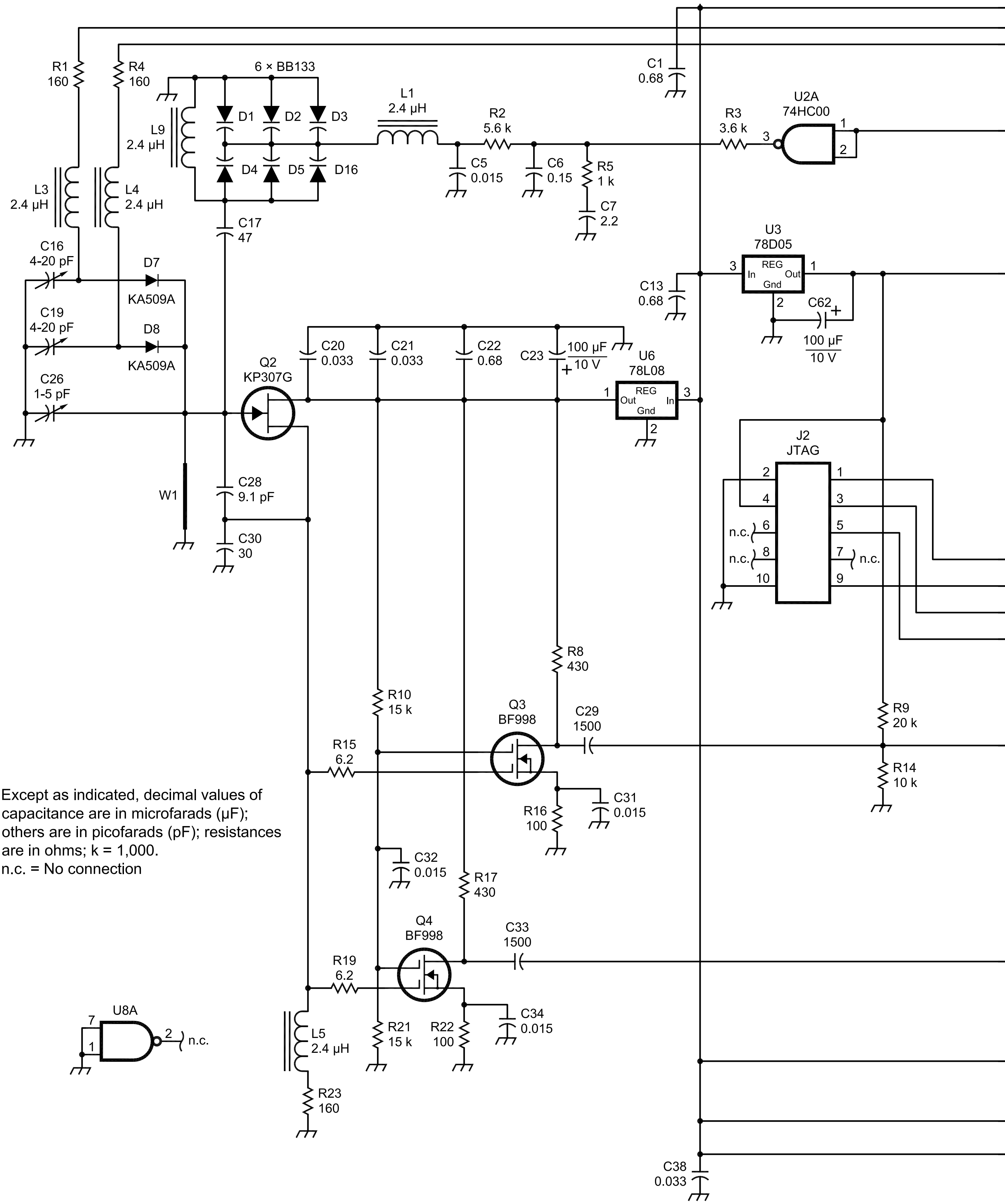
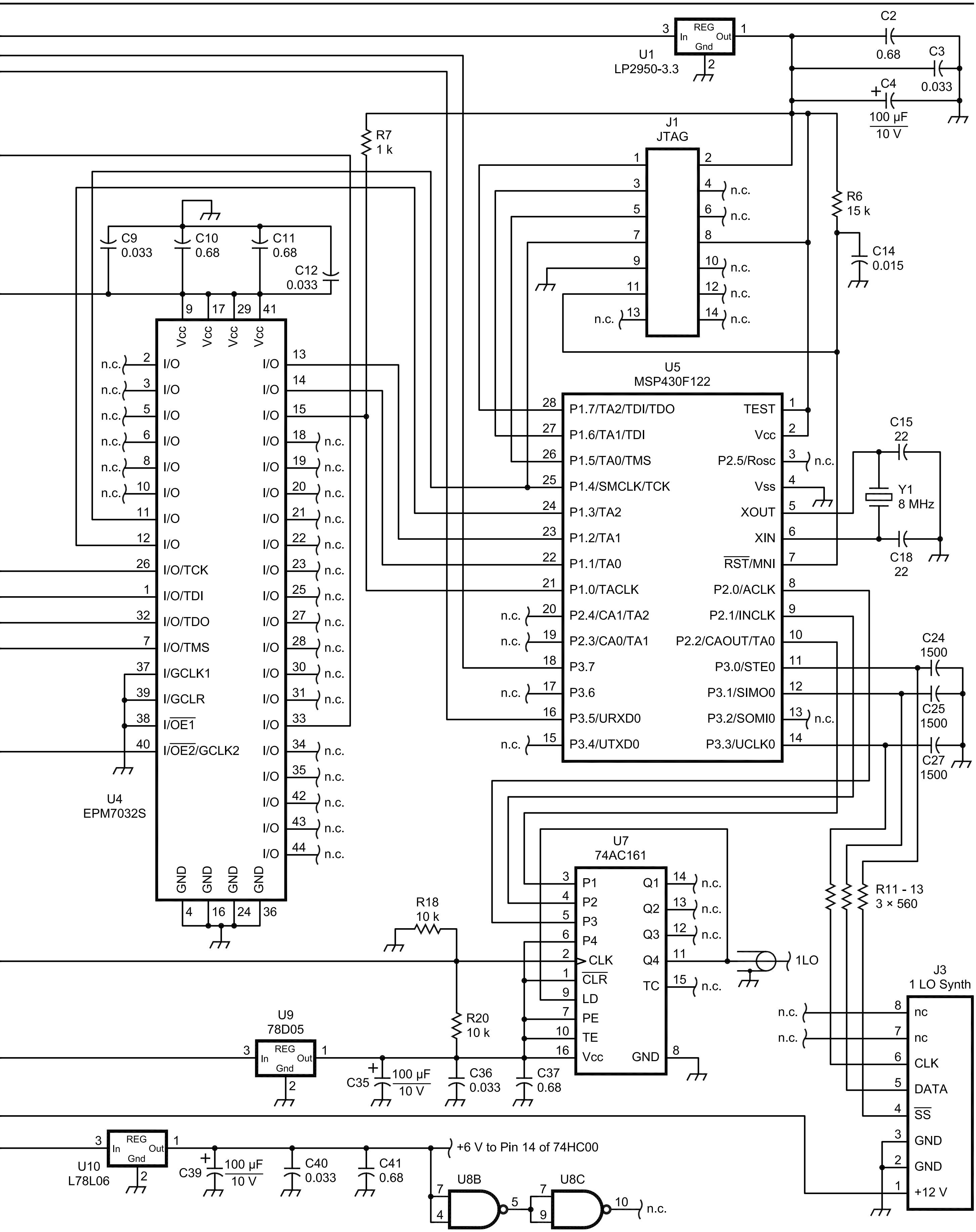


Fig 6—The synthesizer's schematic diagram.


U1
LP2950-3.3
REG
In
Out
Gnd
C2
0.68
C3
0.033
C4
100 µF
10 V
J1
JTAG
R7
1 k
R6
15 k
C14
0.015
C9
0.033
C10
0.68
C11
0.68
C12
0.033
U5
MSP430F122
P1.7/TA2/TDI/TDO
P1.6/TA1/TDI
P1.5/TA0/TMS
P1.4/SMCLK/TCK
P1.3/TA2
P1.2/TA1
P1.1/TA0
P1.0/TACLK
P2.4/CA1/TA2
P2.3/CA0/TA1
P3.7
P3.6
P3.5/URXD0
P3.4/UTXD0
TEST
Vcc
P2.5/Rosc
Vss
XOUT
XIN
RST/MNI
P2.0/ACLK
P2.1/INCLK
P2.2/CAOUT/TA0
P3.0/STE0
P3.1/SIMO0
P3.2/SOMI0
P3.3/UCLK0
C15
22
Y1
8 MHz
C18
22
C24
1500
C25
1500
C27
1500
U4
EPM7032S
I/O/TCK
I/O/TDI
I/O/TDO
I/O/TMS
I/GCLK1
I/GCLR
I/OE1
I/OE2/GCLK2
n.c.
U7
74AC161
P1
P2
P3
P4
CLK
CLR
LD
PE
TE
Vcc
Q1
Q2
Q3
Q4
TC
GND
1LO
R18
10 k
R20
10 k
R11 - 13
3 × 560
J3
1 LO Synth
nc
CLK
DATA
SS
GND
+12 V
U9
78D05
C35
100 µF
10 V
C36
0.033
C37
0.68
U10
L78L06
C39
100 µF
10 V
C40
0.033
C41
0.68
+6 V to Pin 14 of 74HC00
U8B
U8C

design uses 100 kHz).

The CPLD U4 is the smallest device in EPM7000S family. It is configured to contain a prescaler, a reference divider and a dead-zone-free phase frequency detector (PFD). The schematic of the CPLD is shown in Figs 7 and 8. Those diagrams were designed with the E+MAX CAD tools.[9]

The voltage divider (R9, R14) is used to set up the necessary bias for EPM7032S clock input. It has LVCMOS compatible inputs, so it needs approximately 1.7 V.

The prescaler is built using a Johnson counter to satisfy the 100-MHz operating frequency requirements and to obtain square-wave output. Five

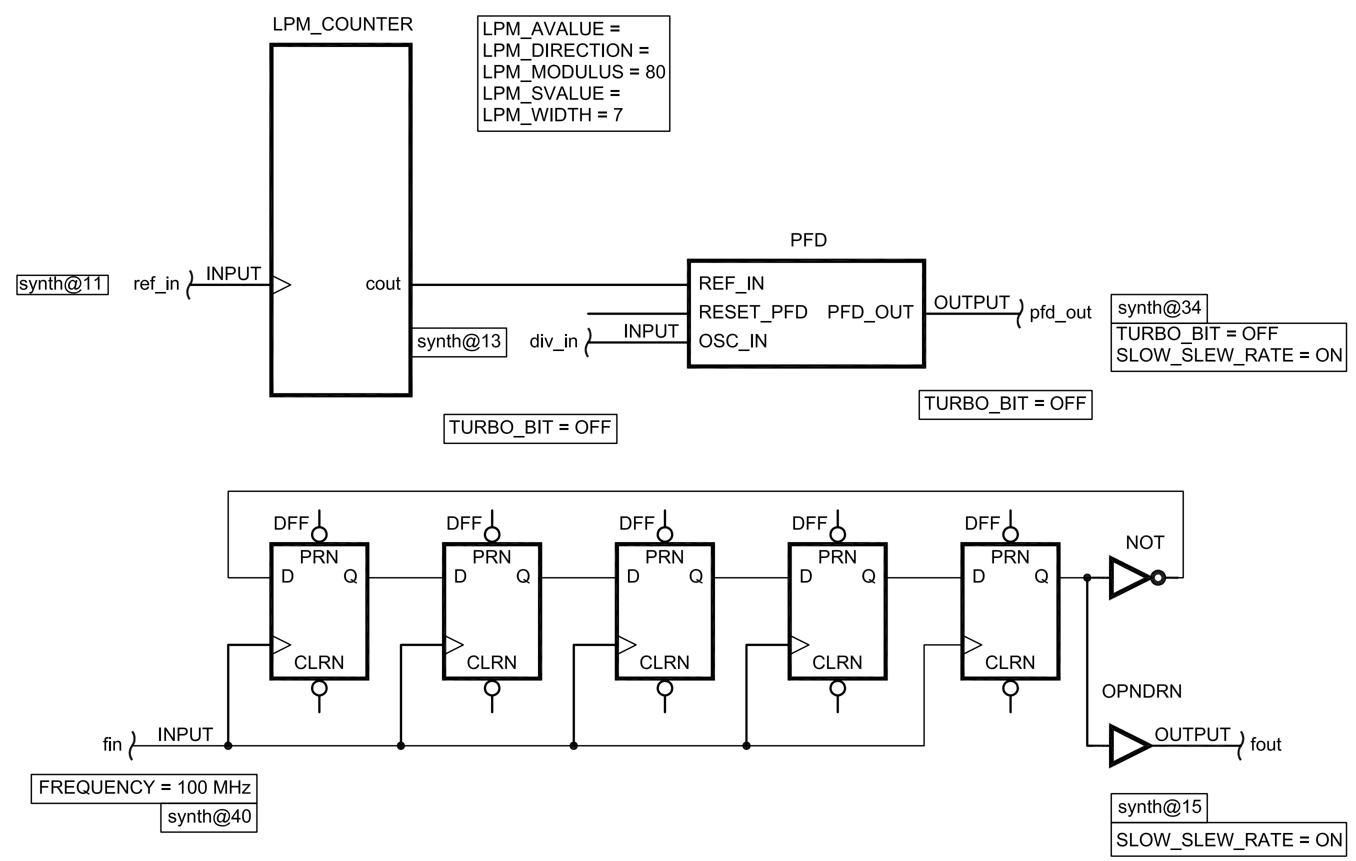


Fig 7—A CPLD functional diagram.

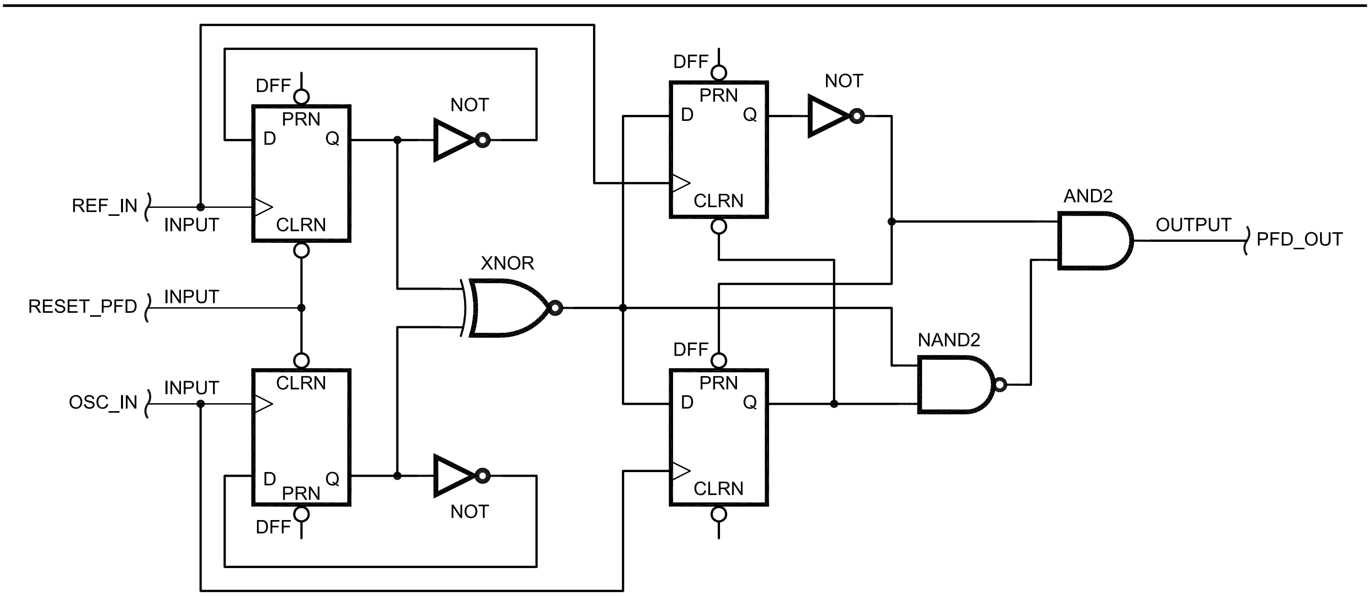


Fig 8—Functional diagram of the dead-zone free PFD implemented in the CPLD.

stages are used for division by 10. The output is fed to the fractional-N divider, implemented in the U5 microcontroller's software. Since U5 is a 3.3 V device, the open-drain output and 1 kΩ pull-up resistor connected to 3.3 V are used.

The reference divider operates at 8 MHz and divides the input frequency by 80 to get a 100-kHz reference, which is fed to the PFD reference input. The microcontroller's oscillator is used as the reference oscillator. To improve its frequency stability, a high-quality vacuum crystal is used here (see Fig 9).

The PFD design is a very important part of the fractional-N synthesizer. The traditional dual D-type flip-flop (DFF) PFD will fail here. As mentioned before, the noise at the output of the fractional divider has a high-pass shape; if we want to have the same noise at the VCO output, we need to have a highly linear PFD. Otherwise, high-frequency noise will be intermodulated down to the low-frequency spectrum. The quad DFF and XOR gate design gives us such an opportunity.

The PFD is composed of two sections (see Fig 8). The first section is the phase detector composed of a dual DFF and the XOR gate, which is activated when the two signals to be compared are close in frequency. The second section is a frequency discriminator composed of a dual DFF and two NAND gates. It overrides the phase-detector section when the inputs have two frequencies far from each other to drive the oscillator frequency toward the reference frequency and put it within range of the phase detector. The detailed description of such a design can be found in the AD9901 datasheet.

The output of the EPM7032S has a 3.8 V logic-one level. This prevents us from connecting it directly to the loop filter. So, I used a 74HC00 gate as an amplifier. It operates at the highest permissible supply voltage (6 V), so I have a 1 to 5.6 V swing at the Varactors. One could say that it would be better to use an op-amp integrator with higher supply voltage here, but this design is simple and does not suffer from the noise and non-linearity of an op amp. By the way, all models of IC-746PRO/756PRO use passive loop filters and even smaller tuning-voltage ranges.

The synthesizer uses a fourth-order, passive loop filter. I have calculated initial values using *MathCAD* and tweaked the values a bit during phase-noise measurements. The corner frequency was chosen to be 400 Hz. The choice was directed by high-frequency noise suppression requirements. The

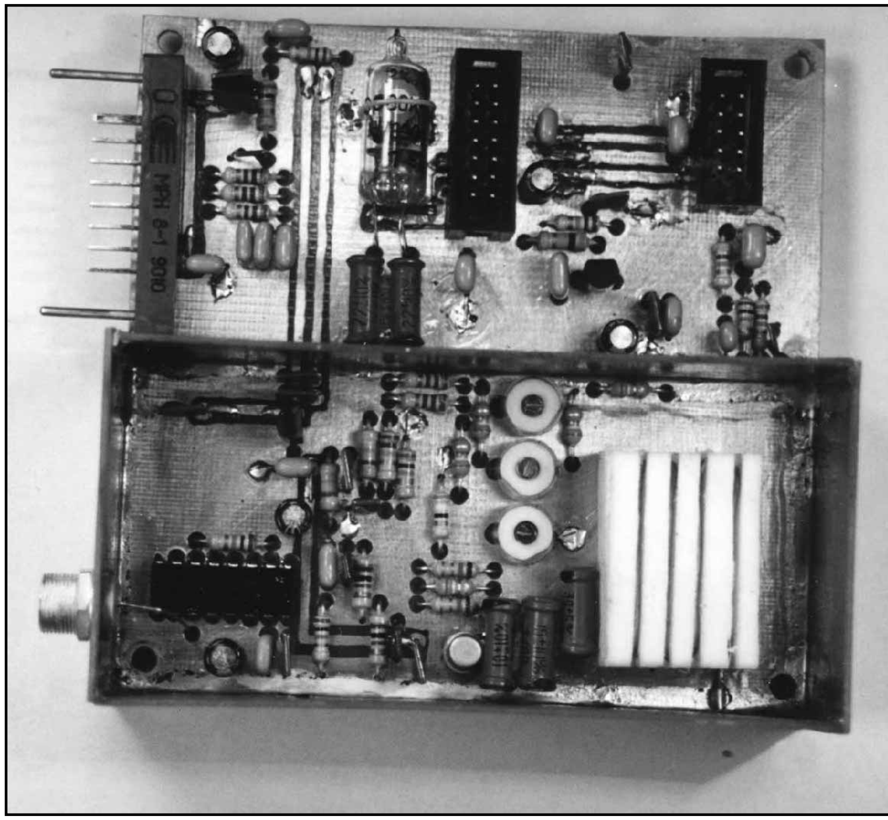


Fig 9—A top view of the synthesizer PC board.

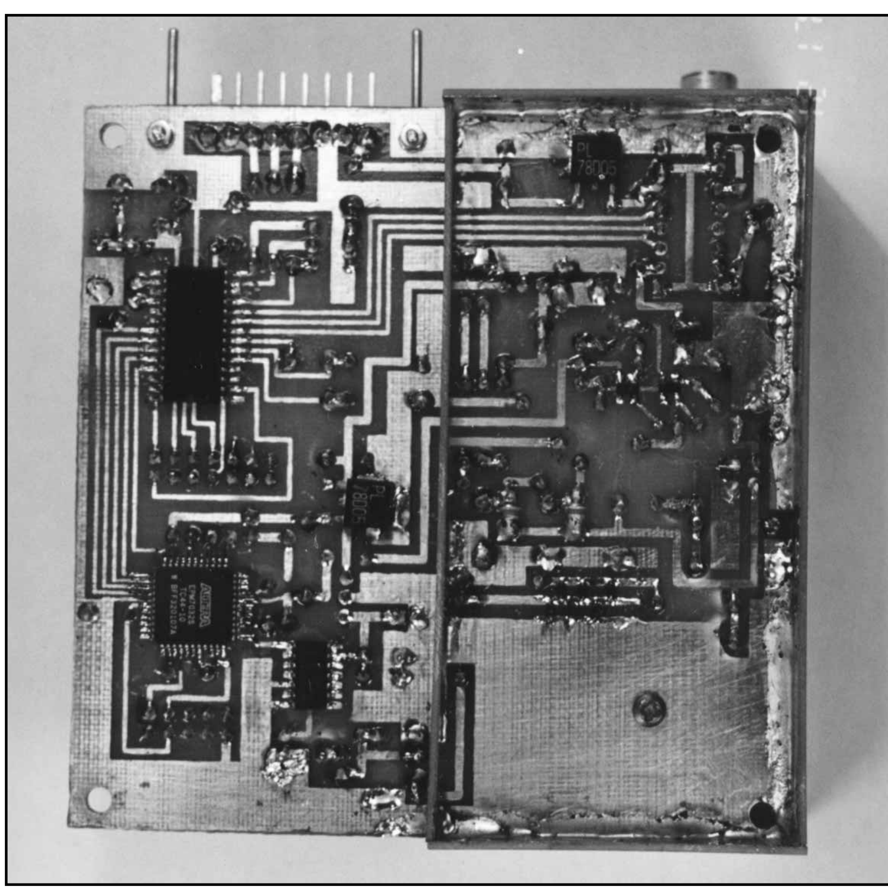


Fig 10—A bottom view of the synthesizer PC board.

Byte number	1	2	3	4	5	6
Description	N lowest byte	N second byte	N third byte	N fourth byte	N highest byte	Control byte

Bit number	7	6	5	4	3	2	1	0
Description	VCO1	0	VCO2	0	0	D2	D1	D0

Fig 11—The control-word format.

simulation of the switching time showed less than 8 ms switching time.

The low-power 16-bit microcontroller, U5, acts as fractional divider and control unit. The software uses the timer module as a divider.[10] When the timer generates a pulse at the output, its interrupt procedure is invoked to calculate the division ratio for the next cycle. This value is loaded into the timer compare register for the next pulse generation. This is repeated 100,000 times per second.

I initially used three 25-bit accumulators (with the 25th bit set to one); but after assembly and testing the synthesizer noise performance, I found some spurs, especially close in. So I decided to try four 32-bit accumulators—they were just what I needed! Fortunately, I need only to make minor changes to software and upload it to microcontroller to try different fractional-N divider configurations.

U5 is also used to control VCO band switching by injecting a 15 mA current through the PIN diodes, setting division ratio of the output divider, U7, and control synthesizer via the SPI bus. The format of the control word is shown in Fig 11.

The next formula gives the relationships of value for the divider (N), output-division ratio (M), output-divider control bits (D) and output frequency (F_{out})[11]:

$$M = 9 - D \qquad \text{(Eq 2)}$$

$$F_{\text{out}} = 10 \cdot F_{\text{ref}} \cdot \frac{256N + 1}{2^{32} \cdot M} \approx 10 \cdot F_{\text{ref}} \cdot \frac{N}{2^{24} \cdot M} \qquad \text{(Eq 3)}$$

$$N = \frac{2^{24} \cdot M \cdot F_{\text{out}}}{10 \cdot F_{\text{ref}}} \qquad \text{(Eq 4)}$$

I think that $\frac{10 \cdot 10^5}{2^{24}} \approx 0.06\,Hz$ VCO step size is more than sufficient, so I decided to discard the low byte of the 32-bit first accumulator input word and set it to 00000001b to ensure longevity (see above). Notice that this value is divided at least by four in the worst-case of my design. The synthesizer uses five voltage regulators to satisfy ICs power supply requirements and get the necessary decoupling between different synthesizer blocks.

Construction

The synthesizer is made on a 95 mm by 95 mm homemade PC board (see Figs 9 and 10). The VCO, isolating amplifiers and output divider are shielded from the other circuitry. The VCO resonator is made from a piece of Teflon (used for mechanical stability) and silver-plated wire (0.5 mm diameter).

Results

The phase noise of the synthesizer was measured using a signal generator (the manufacturer claims it has –140 dBc/Hz phase-noise performance at 100 kHz offset) followed by a two-pole crystal filter. The synthesizer was installed in my homemade DSP transceiver. The test signal was applied to the antenna jack and the DSP block was used

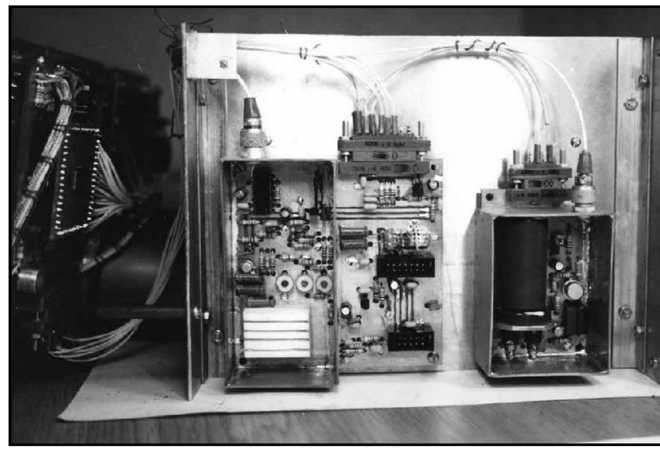


Fig 12—The synthesizer within the transceiver under construction.

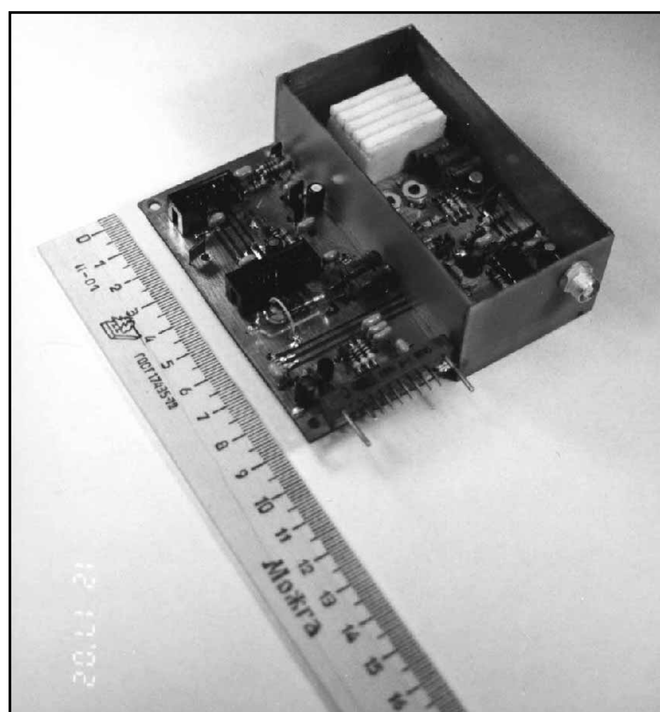


Fig 13—The synthesizer is about 9 cm square.

to measure the noise power in 1000-Hz or 2400-Hz bandwidths (the DSP software is capable of precise power measurements) at the different offsets. Then I corrected results by $-10\log_{10}$ (*Bandwidth*) to get results in dBc/Hz. The results are shown in Table 2. These numbers were obtained with the division by 10 at the output of the synthesizer at 20 m, so I expect degradation of the phase noise by 8 dB at 10 m and improvement by 2.9 dB at 160 m.

I have no spectrum analyzer, but I could not find any significant spurs receiving a test signal. At this time, this synthesizer is installed in my transceiver. I have found superior stability and a much simpler tune-up procedure in comparison with the previous design based on a single-loop synthesizer with interpolation provided by "pulling" the master crystal (a system like the one in the Elecraft K2).[12]

Table 2

Offset (kHz)	*Noise (dBc/Hz)*
2	–120
5	–128
10	–133
20	–136
30	–139
40	–142

Future Work

I have found an interesting patent in the US patent database.[13] It is possible to construct a frequency synthesizer with the reference multiplication factor of 1 using that patent and the fractional-N technique described in this article. So, if we used a low-noise reference oscillator, highly linear PFD and wide loop bandwidth, we would be able to get very good VCO noise suppression. I think that it is possible to build a super synthesizer using such a design. Of course it will not be as simple as the one I have described. It will need some additional analog circuitry and fast, high-density programmable logic ICs. Currently, I am satisfied with the described design and will concentrate on finishing the other software and hardware of my DSP transceiver. You can view the progress of my transceiver at **users.ints.net/skidan/T03DSP**.

Conclusions

I have presented a simple all-digital fractional-N synthesizer. Despite its simplicity and low cost, it has satisfied the first-LO requirements for my HF DSP transceiver. The synthesizer contains no analog circuits, except the VCO, so it is very easy to build and tune up (actually you need just to set up the necessary VCO bands by tuning C26, C16 and C19).

I expect digital fractional-N synthesizers to be widely adopted in amateur and commercial equipment. They offer some advantages over widely used direct digital synthesizers because of their manufacturability, high resolution and the predictability of the resulting noise.

Notes

1V. Manassewich, *Frequency Synthesizers Theory and Design* (New York: John Wiley and Sons, 1976).

2U. Rohde, KA2WEU, "A High-Performance Hybrid Frequency Synthesizer," *QST* Mar 1995, pp 30-38.

3C. Drentea, "Beyond Fractional-N," *QEX* Mar/Apr 2001, pp 18-25; May/Jun 2001, pp 3-9.

4 O. Golovin, "Professionalnie radiopriemnie ustrojstva dekametrovogo diapazona" (Professional receivers for decameter band) in Russian, Moscow, USSR, 1985 (pp 280-282)

5R. Cox, Hewlett-Packard Company, "Frequency Synthesizer," US Patent 3,976, 945, 24, Aug 1976.

6J. Wells, Marconi Instruments, "Frequency Synthesizers," US Patent 4,609,881, 2 Sep 1986.

7C. Hill, "All Digital Fractional-N Synthesizer for High Resolution Phase Locked Loops. Part 2," *Applied Microwave & Wireless* Jan/Feb 1998, pp 38-42.

8V. Drozdov, "Lyubitel'skie KV Transiveri" (Amateur HF Transceivers), in Russian, Moscow, USSR, 1988, pp 26-31.

9E+MAX and other CAD tools for MAX7000S are available from Altera Web-site (**www.altera.com**) free of charge. Also you can find there an easy to build ByteBlaster download cable.

10The software is written in GNU assembler and compiled using GCC compiler. GCC compiler for the msp430 microcontroller series is available free of charge from (**mspgcc.sourceforge.net**).

11In my transceiver, the synthesizer signal is additionally divided by two on the mixer board. Here I give formulas for synthesizer board output.

12You can learn more about Elecraft K2 transceiver (and download manuals with full schematics) on the Elecraft Web-site (**www.elecraft.com**).

13Alexander Roth, Rohde & Schwarz GmbH & Co KG, "Frequency Synthesizer Operating According to the Principle of Fractional Frequency Synthesis," US Patent 5,847,615, 8 Dec 1998.

ARRL 2003 Technical Awards Call for Nominations

ARRL members are encouraged to send nominations to ARRL Headquarters. Please include basic contact information for both you and the nominee. Submit support information along with a nomination letter, including endorsements of ARRL affiliated clubs and League officials. Nominations should thoroughly document the nominee's record of technical service and accomplishments. The nomination form for these awards can be found at **www.arrl.org/ead/award/application.html**.

ARRL Technical Service Award is to be given annually to the licensed radio amateur whose service to the amateur community and/or society at large is of the most exemplary nature within the framework of Amateur Radio technical activities. These include, but are not limited to:

- Leadership or participation in technically oriented organizational affairs at the local or national level.
- Service as an official ARRL technical volunteer.
- Service as a technical advisor to clubs sponsoring classes to obtain or upgrade amateur licenses.

The Technical Service Award winner will receive an engraved plaque. In addition, the winner may request ARRL publications of a value up to $100.

ARRL Technical Innovation Award is granted annually to the licensed radio amateur whose accomplishments and contributions are of the most exemplary nature within the framework of technical research, development and application of new ideas and future systems. These include, but are not limited to:

- Development of higher-speed modems and improved protocols.
- Promotion of personal computers in Amateur Radio applications.
- Activities to increase efficient use of the amateur spectrum.
- Digital voice experimentation.

The Technical Innovation Award winner will receive a cash award of $500 and an engraved plaque.

ARRL Microwave Development Award is given each year to the amateur (individual or group) who conducts research and application of new and refined uses and activity in the amateur microwave bands. This includes adaptation of new modes both in terrestrial formats and satellite techniques.

The Microwave Development Award winner will receive an engraved plaque. In addition, the winner may request ARRL publications of a value up to $100.

Nominate Now!

Send nominations to: ARRL Technical Awards, 225 Main St, Newington, CT 06111. Nominations must be received at Headquarters by March 31, 2004. Send any questions to Headquarters or e-mail **jwolfgang@arrl.org**.

Automatic Signal Classification for Software Defined Radios

How might a radio determine the appropriate demodulator for an incoming signal?

By Frank Brickle, PhD, AB2KT

One of the more intriguing prospects for software radios is the possibility of programming them to act on their own. Some examples:

- Your receiver is configured to scan for activity and then automatically route it to the right place. If it's phone or CW, it's sent to the speakers; if digital, such as PSK31 or MFSK16 or RTTY, the appropriate decoding is applied and the output sent to a window. To transmit in the correct mode, you only have to type, key or start talking.
- You're interested in a new digital mode that doesn't have a lot of users yet. You tell your radio to let you know when somebody comes up in that mode on any of the HF bands, while you go on about your business elsewhere in the shack.
- As you carry on a QSO, a constantly-updated spectroscope of a selected set of HF bands is displayed on your computer screen, with tags indicating the type of activity for each signal as it appears and disappears.
- You have your receiver keep a diary of the times and frequencies a certain type of RFI appears.

6 Kathleen Pl
Bridgewater, NJ 08807
ab2kt@arrl.net

What makes such scenarios possible is automatic signal classification, a technique for examining signals and determining their modulation type and often their protocol, without user intervention. In short, it's a way for a receiver to know to what it's tuned. We could also characterize it as the sensor for an intelligent agent running the radio. Calling as it does on a merging of DSP and statistics, automatic classification is ideally suited to implementation in software radios. The examples mentioned are merely a few ways these capabilities may be further configured in software to perform useful tasks.

Such an introspective receiver is not yet a reality; but much of the underlying technology is in place, mature and open for experimentation. While the full exploitation of automatic classification technology may require software radios, much of the technology can be applied to conventional equipment, especially when computer control is available.

It is not hard to assemble classification systems that are simple and moderately effective. Almost any reasonable approach will beat flipping a coin. In fact, experience suggests that a good first approximation will often get to an 80% hit rate. But that's one error out of five, split between false positives and negatives. There is rarely an easy fix for this. The true work comes in eating away at the remaining 20%.

An Informal Tour

Familiar Classifiers

Conceptually, elementary classifiers have been in use for a long time, although they've heretofore been implemented as analog circuits, and they're usually employed as little more than squelch mechanisms. For example, a traditional squelch uses a discriminator between two crude classes of signal: low-energy (dead air) and high-energy (causal signal). Repeater access control by CTCSS relies on a recognizer for signals carrying tones of specified frequencies. Hobbyist scanners often have a "data-skip" function for avoiding channels that are active but carry something other than analog voice.

The data-skip example differs from the other two in one very important way. Theoretically, its function is to reject excessively coherent signals—those with high energy but low spectral variance—since the signal of interest, speech, exhibits both high energy and high spectral variance. What makes this interesting is that spectral variance is a statistical concept, and its discrimination is based on measurements pertinent to *any* signal; in other words, on *models* of speech and non-speech. This is evident from the way a typical scanner can be fooled by signals that exhibit a fairly coherent spectrum and slow modulation rate, with occasionally rapid on-off keying, and so can look a lot like speech. Some pagers fit this profile very well.

By contrast, the simple energy or CTCSS mechanisms operate directly on measurements of the specific signals of interest. In both cases, what's measured is either the overall energy level or the level at one specific frequency. Whatever the measurement, a squelch technique needs to decide when to let a receiver talk and when to tell it to shut up. Most of the time, turning a knob sets the decision threshold, and the discrimination comes down to a single test: Is that threshold exceeded?

The automatic classifiers we're interested in can be regarded as exalted versions of the data-skip feature. The essential distinguishing property of this family is reliance on signal models. Yet what separates our classifiers is the need to handle a broad array of marginal and unclear cases, many of which will involve weighing uncertain evidence beyond simple yes-or-no decisions. The fundamental tool here is *probability*; and so the core of our classifiers will consist of probability models. Much of the time, the evaluation will require accumulating evidence over many observations, not just choosing among alternatives for a single observation. A realistic classification system generally consists of multiple models applied in parallel to an ongoing stream of input data.

Why Not Brute Force?

The direct approach is always the first one to investigate. In classifying signals that would be: *Process any signal as if it were the one you're looking for.*

If the processing works, bingo. If it fails, declare a miss. When you're trying to classify a signal as one of several possible types, try each one in turn until you get a hit, or they all miss. The conceptual advantage of this approach is that modeling is irrelevant, since each strain of signal-specific processing already embodies what there is to know about its target.

There will be situations when brute force is the sole option, but only as a last resort. The issue is how to know when the processing works, especially under adverse conditions. The only robust, universally applicable method for measuring the quality of processing is to first catalog the output of the processing under an exhaustive set of right and wrong inputs and a full range of signal conditions and underlying data types. This just fobs off the modeling requirement to a different stage of the classification process, and in many instances to a statistical problem that's even more difficult than the one we were trying to avoid by resorting to brute force. So it doesn't really gain us anything, if indeed it's practical at all.

There are other problems. Demodulation and decoding can be computationally costly, and we might not have the time or power for full-up processing in a given application. The processing might be only marginally effective, as with manual CW or noisy speech. We might not actually possess the necessary software to perform the full processing, as with a proprietary modem. Worst, we might be interested in classifying a signal that exhibits structure but doesn't actually bear any information, as with various kinds of stereotypic EMI or RFI, so there is no processed output to consider. Our methods need to be able to discriminate among these cases nevertheless.

Our Holy Grail is the *universal classifier*—a clear notion of what the design would be, whether we attain it at the moment or not. What's implied is the need to always start with the most general setting and with the fewest assumptions about the kinds of inputs and outputs the classifier needs to handle. Parenthetically, this goal also implies that we would ideally want to be capturing data in quadrature, prior to any sort of demodulation.

A One-Note Tune

Fig 1 is the block diagram of a pointedly naive automatic classifier system, meant to monitor a single frequency and raise an alert when a signal of one specific type appears. Other signals are rejected. This is a simplistic case, but it embodies the overall shape of any system.

Our immediate concern is an intuitive overview of classifier design. We will come back to the formal details later.

It's important to keep in mind that a classification *system* comprises a number of elements, only one of which is properly designated a classification *algorithm*. At this stage, we are noncommittal about exactly where and how the components—the radio, the programs and the alert—are realized in terms of hardware and software. In any case, the left half of the diagram is mostly signal process-

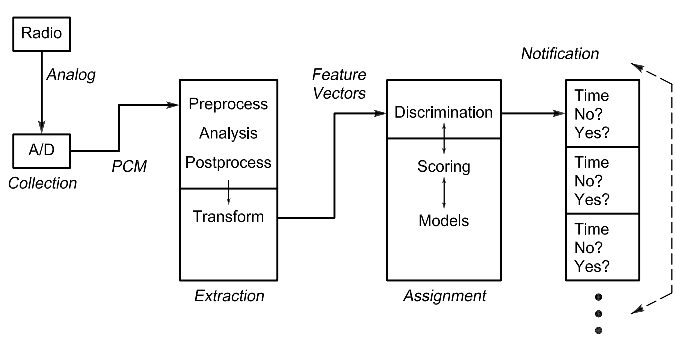


Fig 1—A block diagram of a signal classification system.

ing and the right half implements the classification algorithm proper. The overall flow is straightforward, as shown in Fig 1.

Collection: The audio signal from the radio is converted by the A/D to buffers (frames) of PCM.

Extraction: The incoming PCM frames, with some pre- and post-processing, are rendered down by analysis and post-trans-formed into sets of characteristic attributes, the feature vectors.

Assignment: The attributes are compared with the target signal by the discrimination process, which scores them against a stored model. The results are written out along with timestamps.

Notification: The time-stamped results are monitored. When a there is a close match to the target, an alert is triggered.

From the standpoint of algorithm design, these four components represent a classic chain of reader-writer coroutines.[1]

A Real Signal

Let's fill in some details. The target signal is MFSK16 on 10.147 MHz USB. Receiver audio is patched to a soundcard on a PC, so the digital samples convey a real rather than a complex signal. The signal is sampled at 11.025 kHz with 16-bit resolution. The receiver passband is about 4 kHz wide.

For simplicity we'll assume that only one signal at a time will appear in the passband. This is asking a bit much from a conventional system, but is quite feasible in a software receiver with wide A/D bandwidth, a preliminary layer of energy detection and selective channel downconversion. For the moment, we'll just assume the channel isn't very busy.

Fig 2 shows the long-term power spectrum of a typical MFSK16 signal. Fig 3 is an enlarged view showing the spectral features of the signal in more detail. This example was generated off-line using a sound card digital program, so the quality is much better than would be expected from an actual HF signal.

An idealized MFSK16 spectrum is shown in Fig 4, and the sample spectrum in Fig 3 certainly looks a lot like it. Such a representative or model spectrum is often called a *signature* or *template*.

The task of a naive recognizer for MFSK16 can be stated in a few words: *Wait for a signal that looks like the MFSK16 signature.* Put in the most colloquial terms, you test a spectrum for similarity by laying it on top of the signature and seeing how well they line up. That is all the block diagram is meant to show: Frame by frame, the incoming signal is lined up with the signature, and when there's a good match, a hit is announced. Notice that the process is running continuously. Even dead air is a signal in this context, and it's not often truly dead.

In the block diagram, the spectral frames are produced in the analysis section of the extraction phase. They are probably—but not necessarily—implemented with an FFT and some ancillary computations. In the assignment phase, the signature is the model, the process of lining up a spectral frame with the signature is the scoring procedure and the decision as to whether the match is close enough is the discrimination component. Intuitively, we might expect the scoring to be realized by some kind of vector computation like Euclidean distance, or the inner product, or a vector cosine. The notification phase might smooth out the results by watching the discrimination output and looking for agreement among some number of the most recent reports.

Problems Right from the Start

Of course, none of this is quite as straightforward as we've made it seem. Previously, we attached a lot of importance to the idea of probability models and scores, and that principle hasn't emerged yet in this sketch. Before addressing that point, we need to consider some of the complications that arise even in this superficial description.

The first issue is already evident in Fig 2. The receiver passband is considerably wider than the target signal, and the target signal might show up anywhere in it. This will require hunting for the spectrum within the passband. The task is made a little simpler by stipulating that only a single signal will appear in the passband at any one time. Hunting can incur a fair computational burden, as it potentially involves computing a full match at multiple positions in the passband. It might be desirable to attack all these com-

[1] This has a number of interesting implications, the most important being that they are asynchronous in principle. *There is no global timing.* The practical consequence is that the individual phases are absolutely indifferent to whether their data sources and sinks are files or other components, or indeed even on the same landmass. They work perfectly as Unix processes using standard input and output, and thus would be capable of running off-line stand-alone, in a pipeline, or via sockets using *netpipes*.

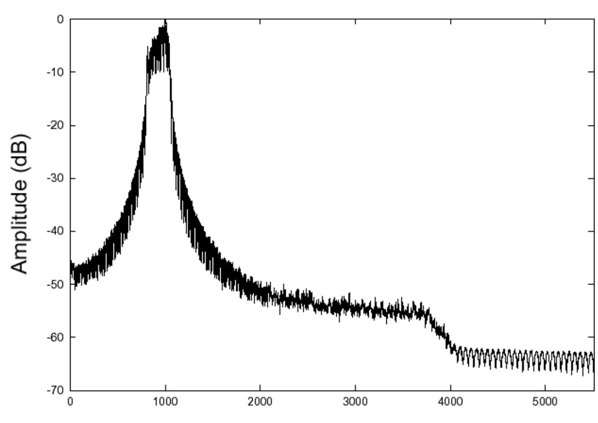


Fig 2—Long-term power spectrum of an MFSK16 signal.

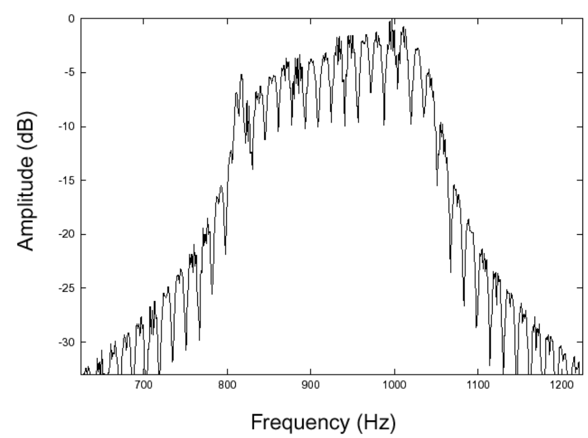


Fig 3—An enlarged view of Fig 2.

plications at once through the use of FFT-based cross-correlations as the matching technique, thus gaining both computational efficiency and position independence within the passband. Yet that approach introduces other, possibly more severe problems down the line, since, as mentioned before, the resulting correlation coefficients themselves will then probably require modeling and scoring to yield sensible answers for a full range of possible inputs.

A more serious complication is illustrated in Fig 5. This figure shows a patch of the spectral surface of the signal used to produce Figs 2 and 3. It represents a sequence of short-term power spectral estimates of the signal, truncated here both in time and in frequency for visual clarity. Each of the short-term estimates conveys the localized frequency content of a single frame, and thus represents what will be matched against the signature. None of the short-term spectra looks altogether much like the ideal, which is merely another way of remarking that the signal is indeed modulated.

To summarize, the naive method of signature matching is inadequate because *any* observations we make on an input stream of genuine data simply aren't going to look very much like the signature. However we choose to measure similarity, the short-term spectra will achieve only partial or poor matches. What's more, we are using great amount of data (many spectral points in both the observations and the model) when we suspect that only a small number of them, basically inflection points, are actually significant.

Fixes

One remedy is to combine the short-term spectra into longer-term estimates and compact them by merging neighboring frequency bins. There is a lot to recommend this idea, and in fact, averaging is one of the more powerful tools in the classification repertoire. Unfortunately it has no effect on this fundamental property of modulated signals: The frequency content depends on the *modulating* signal. The data used for Figs 2, 3 and 5 were deliberately concocted to be representative of an MFSK16 signature. There is no reason to assume that an arbitrary MFSK16 transmission will embody such representative underlying information, or will be long enough to produce a sufficiently sharp long-term estimate. Furthermore, there will be scenarios in which the differences between successive spectral frames, as expressed in the spectral variance, will play heavily in the discrimination of, say, speech signals, or similar but not identical signal types like MFSK8. So a technique (like time-averaging) that is intended precisely to reduce spectral variance could wind up working against us.

What's required is to take as observations the smallest distinguishable set of attributes, the *features*. We can derive the model by measuring the values of these features over a wide variety of inputs and computing the corresponding probability distributions. In this way, the uncertainty and variability of the observations are incorporated in the model as the parameters of the probability distributions. For the moment, we will dodge the question of how to discover the right features, except to mention that this is a classification procedure too, but in an *unsupervised* or *blind* setting.

Fundamentally, the score for a candidate signal is just how likely it is based on the probability of its observed features under the model. The process of estimating the parameters of the model is referred to as *training*, and the body of sample material used to train is called the *corpus*. The individual members of the corpus are *exemplars*. In the training process the features are extracted from the exemplars by exactly the procedures used in the scoring process. The modeling of the features can be said to predict the features of a new signal of the correct type.

The most desirable situation is when the exemplars are drawn from live collection using the components with which the classifier will be implemented, preferably in quadrature. Largely, this is because there are many potential complications that are circumvented by being folded into the model and the training process. For example:

- *Noise:* The signal in Figs 2 and 3 is very clean, but the features corresponding to the MFSK16 tones can be obscured, in this view, by an increase in noise level of only 5 dB. In situations where noise is basically random it can be viewed additively. The noise contribution is absorbed in the variance estimates of the probability distributions on the features. Where the noise is structured, as with birdies or heterodynes, we can expect to have mitigated them prior to the feature extraction step, in the preprocessing phase.
- *Channel characteristics:* As is evident in Fig 2, the input passband starts to roll off at about 3.8 kHz. A signal close to the higher band edge can be subject to significant amplitude (and phase) distortion. If the model has been trained on signals placed in a variety of

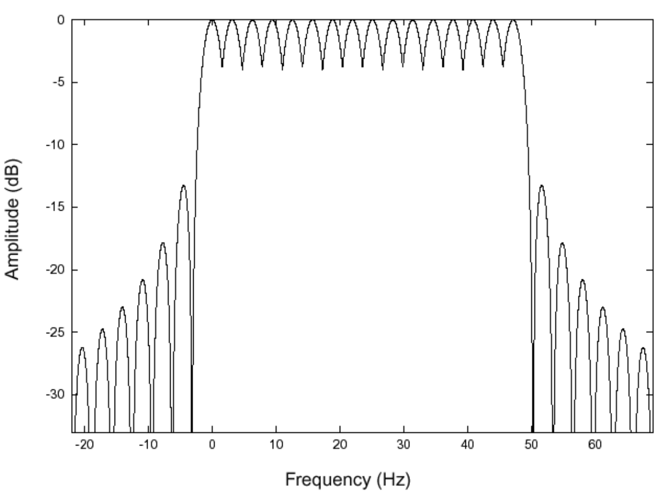


Fig 4—Ideal MFSK16 spectrum.

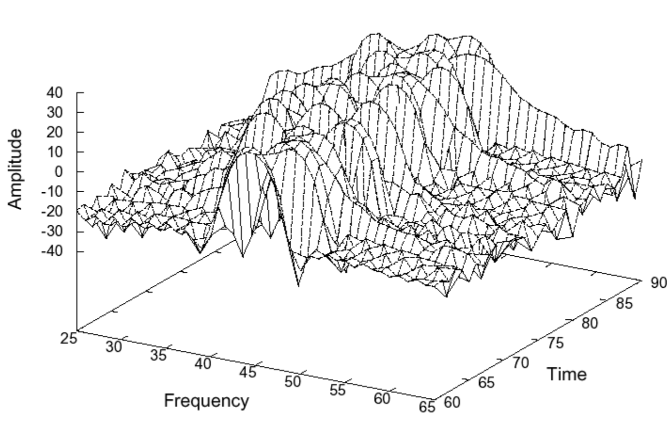


Fig 5—Excerpt of MFSK short-term power spectrum surface.

positions in the passband, that distortion is absorbed into the probability distributions as well, with estimation artifacts due to the size and resolution of the spectral analysis procedure.
- *Confusion and misclassification cost:* In the real world, HF anyway, many cases are difficult to discriminate. Probability scoring gives an answer on *all* possible candidates, hence a quantitative handle on how right we think we are. This is essential information for evaluating *misclassification cost*, which comes down to how much importance we attach to a given class. We need to know this to decide how much work we can afford to spend on a given classification, which translates directly into how complicated we can afford to be.

So a number of difficulties are addressed at once by adopting probability models of spectral features. However, it must be stressed that by incorporating uncertainty into our models, we have given up on quick answers. Just as the model parameters were derived from observations on a lot of data, so the discrimination power of the model now has to be teased out of multiple measurements of a signal over time. We've also abandoned a direct way to update the classifier. The entire system will probably need to be retrained if a signal class is added or removed. And we've given away some portability and flexibility by asking that the exemplars be collected from our own radios. These are problems in principle and there are heuristic ways around them, but they are beyond our scope right now.

A Universal Classifier?

It's reasonable to ask the limitations of this general classifier design, particularly with respect to the kinds of signals it's capable of recognizing. The answer is that in principle—except for analog speech—pretty much any kind of signal can be recognized, including structured noise from EMI and RFI. Most current limitations are practical and imposed by hardware. For example, spread-spectrum signals are difficult to process with affordable A/D converters on account of their wide spectrum usage. Yet as full-IF-bandwidth A/D converters become economical, and as software radio technology migrates more from DSP chips into general-purpose computers, such limitations begin to fall away. Other problems recede as well, as capabilities like dynamic resource allocation and adaptive channelization over a wide bandwidth become commonplace.

Better hardware does not address the problem of speech detection, however. It's not that the kinds of detectors we're describing are incapable of finding speech. Once again it's because they're so easily fooled by signals that look like speech to them but aren't, especially in noise. What's hard about speech is that it is so *variable*, on top of exhibiting considerable but inconsistent redundancy of spectral properties. This is evident from the proliferation of techniques for compressing and coding speech for digital transmission, which can differ significantly in what they regard as essential or dispensable information. It is an amusing exercise to listen to the output of a speech decoder working on input made up from credible but randomized parameters. Very quickly you get a clear notion of what that particular coder thinks is "average" speech. Doing this exercise for a number of coders gives a similarly clear picture of how much they diverge from one another in their concepts of what uniquely characterizes human utterances.

What's missing in these examples, of course, and what we humans put into the signals, is *patterning* or *sequencing* in time of the short-term frequency properties. This patterning in time is the "markovity" of the acoustic speech signal, and it represents the influence of recent past events on the most current ones. Endowing a probability model with this kind of memory is not trivial. You can get away with treating spectral observations of most signals as independent; they're processed sequentially because that's how they arrive, one at a time. But with speech, higher-order markovity—memory of events many steps in the past—is the critical determinant, and there is no alternative but to place it at the forefront.

This can be accomplished in a couple of different ways, either enumerate: (1) several not-quite-orthogonal models functioning over multiple time scales, or (2) a considerably more sophisticated model and scoring technique, one that takes into account the higher-order markovity of the signal. But either way, the processing is algorithmically and computationally a lot more demanding. This is a large topic that ought to be taken up independently.

Intermission

We're about to plunge into some more theoretical aspects of classification and signal processing, so we'll pause here to recall the highlights so far. The important points are:

1. Signal classification proceeds by comparing observations of an unknown signal to models of one or more known signals.
2. The observations are snapshots of the frequency content of an ongoing signal, and the models also are built from streams of such observations.
3. It's necessary to model both wanted and unwanted signals.
4. The comparison is probability-based.
5. The classification decision depends on accumulating evidence from a sequence of observations.
6. If possible, observations contributing to the models should be made on actual signals in their native environments.

Better yet, here's a summary metaphor. If you use any of the modern sound card digital modes, you're doubtless very familiar with the waterfall spectral display that most of them use. On one of those displays, each thin horizontal slice represents a brief snapshot of the frequency content of the radio passband, and a signal shows up as a variegated vertical stripe that speckles and bulges as it passes by. The job of a classifier is to predict the flow of speckles and bulges–not just a single horizontal slice, but a whole vertical band. It bases its predictions on having looked at a lot of stripes of the right kind, but just as importantly, having looked at a lot of the wrong kind as well. Note that from this point of view, what a classifier is doing is little different from predicting what will be coming up in a signal from the part of it that's already been seen. (The best predictor for tomorrow is today.)

That's quite a bit already. At this point we turn our attention to the underlying ideas.

A Classification Framework

This section is intended as an orientation to classification theory. The focus is on larger aspects of the theory that have direct impact on what kinds of signal processing techniques will be useful for classification. A comprehensive treatment can be found in several of the excellent references.

A Taxonomy of Classifiers

Classification is a procedure whereby a set of *instances* is assigned to *classes* or *types*, based on *features* or *attributes*. In formal shorthand, a classifier maps vectors to integers where the vectors are sets of features obtained from *observations* or *measurements* of the instances. The integers are indices of the set of possible classes or types.

Classification has been approached from a few different starting points:

- Statistics, based on probability models,
- Artificial neural systems, or neural nets, that attempt to model or simulate processes underlying human performance, and
- Machine learning, an attempt to model or mimic deliberate, cognitive concepts and processes employed by people when doing the identification.

The statistical and neural approaches are essentially quantitative, falling into three components:

1. The underlying distribution of the population of interest—that is, the prior or marginal probabilities of the classes.
2. The discrimination criterion—that is, the set of features and the rules or procedures that use them, to distinguish the classes.
3. The misclassification cost—that is, the importance attached to each class.

Machine learning by contrast is less purely quantitative, since it can rely in principle on combinations of quantitative and nominal observations. Thus the observations in machine learning are more precisely characterized as multiples rather than vectors. Machine learning algorithms can be harder to factor neatly into components, since they frequently embody recursive components and therefore are better abstracted as directed graphs rather than trees.

In practice, a classification *system,* as distinct from a classification *algorithm,* is rarely based on one of these approaches alone. The hypothetical system in the informal discussion is a good example, as it employs quantitative discrimination in the assignment stage, but relies on a certain degree of empirically based logic in the notification phase to produce a final answer.

Components of a Statistical Classifier

For purposes of the present discussion, we will treat classes, types and models as equivalent, although this is something of an oversimplification. We denote the classes or models as m_i, i = 1, 2…M and their unconditional or prior probabilities as $P(m_i)$. In a situation where misclassification costs are all equal or unknown we might apply the prior probabilities alone in the following classification rule:

$$\hat{m} = \arg\max_i P(m_i) \qquad \text{(Eq 1)}$$

where $\hat{m}$ denotes the best choice of class. So the rule means:

Always pick the class with the highest prior probability.

In other words, always guess the class that occurs most frequently.

This rule requires only elementary training from the corpus—basically, just noting how often each class occurs—and incorporates no observations at all about the exemplars in the corpus or any new instances to be assigned. Nevertheless it's the best bet, at odds $P(m_i)/(1 - P(m_i))$. It is sometimes referred to as the *default rule*. Notice that it only applies in the absence of significant misclassification costs.

Yet we do have measurements of the exemplars and new instances that we denote as observations, O, and we can estimate both the marginal distribution $P(O)$ and the conditional distributions $P(O\,|\,m_i)$ of the observations themselves in the corpus. With this information, we can do a little better:

$$\hat{m} = \arg\max_i P(O\,|\,m_i)P(m_i) \qquad \text{(Eq 2)}$$

which means:

Pick the class with the highest posterior probability of the observation, or *pick the class that makes the observation most likely.*

Further, applying Bayes' Rule, we have:

$$P(m_i\,|\,O) = \frac{P(O\,|\,m_i)P(m_i)}{P(O)} \qquad \text{(Eq 3)}$$

and thus we can stipulate:

Pick the class with the highest posterior probability, given the observation.

or

Pick the class with the strongest evidence in its favor, based on the observation.

One form of the optimal discriminator for an observation O and two classes m_i and m_j is the *likelihood ratio* $P(O\,|\,m_i)\,/\,P(O\,|\,m_j)$, which is interpreted as the weight of evidence for m_i against m_j. Once again, this holds as a maximization problem if we ignore misclassification cost.

Now, if misclassification costs do need to be considered, we must treat the problem in terms of minimum risk rather than maximum likelihood. We denote the cost or risk of misclassifying class m_i as m_j by $c(i, j)$. The cost is not generally symmetric—$c(i, j) \neq c(j, i)$ except when $i = j$.

Going back to the default rule, where every new instance is assigned to the most probable class, we can compute the expected misclassification cost of assigning everything to the same class. Let the default class be m_ϕ, the cost of each classification of m_i to m_ϕ, be $c(i, \phi)$ and the expected cost of the default decision be C_ϕ, . Then:

$$C_\phi = \sum_i c(i,\phi)P(m_i) \qquad \text{(Eq 4)}$$

which merely weights the cost by the prior probability of each class, and we choose ϕ to minimize C_ϕ, . In the case where all misclassifications have the same cost—the most common situation in signal classification—we would have constant $c(i, j) = \kappa$ for $i \neq j$ and $c(i, j) = 0$ for $i = j$. The expected cost is then:

$$C_\phi = \sum_i c(i,\phi)P(m_i) = \sum_{i\neq\phi} \kappa P(m_i) = \kappa \sum_{i\neq\phi} P(m_i) = \kappa(1 - P(m_\phi)) \qquad \text{(Eq 5)}$$

so the minimum cost assignment is precisely the maximum prior probability choice of m_ϕ, , where maximum $P(m_\phi,)$ minimizes expected cost $\kappa(1 - P(m_\phi,))$.

The same holds true for equal misclassification costs and posterior probabilities. Yet with unequal costs, an assignment to class m_i must meet the condition:

$$c(i,j)P(O\,|\,m_j)P(m_j) < c(j,i)P(O\,|\,m_i)P(m_i) \qquad \text{(Eq 6)}$$

which means, an i-for-j false negative is less risky than a j-for-i false positive, weighted on each side by the likelihood of the respective wrong case. Further:

$$\frac{P(m_i)\,|\,O}{P(m_j)\,|\,O} > \frac{c(i,j)}{c(j,i)} \qquad \text{(Eq 7)}$$

which expresses the quantitative constraint that the evidence for choosing i over j must strictly outweigh the expected relative risk of mistaking i for j. Trivially, when misclassification costs are equal, the likelihood ratio must be greater than 1. If i is riskier than j, the threshold goes above 1, meaning stronger evidence in favor of i will be needed.

Sequential Testing

As mentioned, the misclassification cost of individual observations is not often an issue with signals. In large part that's because the information about a signal arrives piecemeal over time. The classification procedure is about absorbing and accumulating the evidence as it unfolds. At some point, enough will have been gathered to make a decision. It is here that classification risk plays a critical role, again as previously mentioned, in a trade-off with the cost of making and scoring the observations themselves.

The essential idea is this: As information about an unfolding signal emerges, the likelihood ratios computed from the observations are monitored and accumulated. We would like to have a decision as early as possible. So the criterion for a decision, as well as the criterion for when to stop and return an answer, is based on this rule:

Stop and return an answer when the cost of taking an additional observation exceeds the misclassification cost.

Often, the cost of an additional observation is simply measured in time lost waiting for an answer. So, for example, when processing a sequence of observations $O = O_t, O_{t+1}, \ldots O_{t+k}$, and the cost of taking an additional observation is $t + k + 1$, you just stop when $t + k + 1 > c$, that is, when you've counted off enough observation steps. One consequence of this rule is that in a multi-way classifier with unequal risks, different assignments may require different lengths of time to reach a stopping point.

The general heading for this topic is *sequential probability ratio testing*, and the pertinent theory is treated by *optimal stopping*. You will probably have noticed the resemblance of this process to Viterbi decoding or any other kind of dynamic programming technique for that matter. This is no accident, insofar as they're all Markov processes under the skin, and traversals of weighted digraphs below that. To go further here would be more than we've bargained for in this discussion. Nonetheless, we should at least note again that the multiple observations implied by general sequential testing also yield an important side effect when dealing with signals specifically, in that taking multiple measurements is the fundamental technique in accounting for the variance of spectral estimates.

Where the Probabilities Go

Let's return to the block diagram in Fig 1 and see where all these computations plug in. A lot of preparatory work has gone on before the block diagram is ready to run: A suitable training corpus has been analyzed to get estimates of the prior and conditional distributions $P(m_i)$, $P(O)$ and $P(O\,|\,m_i)$. These estimates all live in the bottom scoring/modeling part of the assignment component. As feature vectors pass into the assignment box, they assume the place of new observations O and are used by the discrimination process to update accumulating model probabilities $P(m_i\,|\,O)$. The assignment process is writing out a sequence of reports essentially representing the state of accumulation. Until there is sufficient evidence for a positive identification, the reports indicate no activity. At the time the evidence tips positive, the reporting changes, and the notification process is cued to raise an alert. This is the point at which the selection $\hat{m} = \text{argmax } P(O\,|\,m_i)P(m_i)$ or the equivalent cost minimization is finalized. If the process keeps running, the evidence should eventually go against the assignment, indicating the signal has gone down.

We've skated right past the issue that the observations will be vector-valued in nearly any interesting problem. We betrayed our hand early by calling classification a procedure for mapping vectors to integers. To round out the theoretical picture we need to address where multivariate probabilities come from. The practical handling of multivariate probability models is well developed, fortunately, and there is a lot of software to assist in that task. Some of the best of it is free. See the sidebar "Doing It Yourself in Software."

Doing It Yourself in Software

All this software is available free for download on the Internet, most of it GPL, some under noncommercial license, some totally unrestricted—generally targeted at *Linux*/BSD systems.

- *LNKnet:* This mother of all classification utilities is available from MIT. It has comprehensive facilities for testing and constructing virtually every known classification technique, with both graphical and command-line interfaces. With copious documentation and examples, it will generate C code for inclusion in a user program.
- *R and octave:* GNU packages implementing most of the *S* statistical package (*R*) and M*a*tLab (*octave*). It has extensive user-contributed add-on packages for many sophisticated and state-of-the-art statistical and classification techniques. It's available from Free Software Foundation mirrors.
- *snd:* Interactive sound-file editor with extensive multitrack recording and programmable signal- processing support, by Bill Schottstaedt, available from Stanford CCRMA.
- *gnuradio:* GPL software radio project.
- *psp:* This flexible spectrum analysis utility by Frank Brickle, AB2KT, can produce ASCII or binary files of power spectrum frames, or long-term spectrum estimates.
- *GPL:* Useful to produce measurements and observations for experimentation with *LNKnet*, *R*, *octave*.

From Power Spectra to Feature Vectors

Most readers will be acquainted with the basics of power-spectrum estimation using fast Fourier transforms or auto-regressive methods; and there are references galore, so we won't dwell on them here. Our concern is with how short-term power spectrum estimates of individual sample buffers, the frames, get turned into feature vectors to be used as multivariate observations, and how the multivariate observations are scored.

The Centroid and Dispersion of a Waterfall: Recall the waterfall spectrum display familiar from programs like *DigiPan* and *MixW*, where each horizontal slice represents a short-term power spectral estimate, and the display moves vertically representing the passage of time. We're concerned with a two-dimensional table $h(t, n)$ that is effectively a frozen chunk of a waterfall display; the table entries correspond to the pixel intensities of the waterfall. So the columns of h (indexed by $n = 1.N$) represent the frequency bins of the short-term power spectra, and the rows (indexed by $t = 1 \ldots T$) correspond to the timed sequence of spectral measurements. In other words, the rows of $h(t, n)$ are vector-valued measurements of the original input signal.

Each vector-valued element can be viewed as a point in N-dimensional measurement space; and if you could plot the points, you would see that they tend to fall into one or several clumps or *clusters*, indicating similarity of the vectors. The regions in which the clusters occur can be characterized by a central location, the *centroid*, and by how dense or sparse the cluster is—the *dispersion*. Any new point—that is, any new measurement vector—can be tested to see which centroid it falls closest to, and how the distance to that centroid compares with the distances of all the other points in the cluster to the centroid. That comparison lets you judge how strongly the new point belongs in the cluster.

More precisely, a new point falls at a distance from every centroid, and therefore can be regarded as belonging to each centroid's cluster, with a strength inversely proportional to the distance. We interpret the strength of membership in each cluster as the probability of belonging to the cluster. That probability is inversely proportional to the distance from the centroid.

Since the whole bundle of points can be regarded as a single cluster, a certain amount of judgement is required to decide how many centroids (hence how many subclusters) we think there really are, which in turn controls how tight the clusters are, and how to assign ambiguous cases.

The *mean vector* gives a quantitative description of the centroid, and the *covariance matrix* characterizes the dispersion. For our table *h(t, n)*, the mean vector is:

$$\mu = \mu_i = \frac{1}{T}\sum_t h(t,i) \qquad \text{(Eq 8)}$$

with components that are simply the means of each of the columns of *h*. The covariance of two columns *i* and *j* of *h* is given by:

$$\operatorname{cov}(i,j) = \frac{1}{T}\sum (h(t,i)-\mu_i)(h(t,j)-\mu_j) \qquad \text{(Eq 9)}$$

Notice that the covariance is a scaled version of the correlation between the two columns. The *variance* σ_i^2 of a column *i* alone is *var(i)*= σ_i^2= *cov(i, i),* with *standard deviation* σ_i, so the correlation ρ_{ij} is:

$$\rho_{ij} = corr(i,j) = \frac{\operatorname{cov}(i,j)}{\sigma_i \sigma_j} \qquad \text{(Eq 10)}$$

The $N{\times}N$ covariance matrix *V* assembles the covariances of all the columns of *h*:

$$\begin{matrix} \operatorname{cov}(1,1) & \operatorname{cov}(1,2) & \ldots & \operatorname{cov}(1,N) \\ \operatorname{cov}(2,1) & \operatorname{cov}(2,2) & \ldots & \operatorname{cov}(2,N) \\ \ldots & \ldots & \ldots & \ldots \\ \operatorname{cov}(N,1) & \operatorname{cov}(N,2) & \ldots & \operatorname{cov}(N,N) \end{matrix} \qquad \text{(Eq 11)}$$

At Long Last, the Punchline: The probability of a multivariate measurement point $\bar{u} = u(n)$ is given by:

$$P(\bar{u}) = (2\rho)^{-\rho/2} \exp\left[-\frac{1}{2}(\bar{u}-\mu)^T \mathbf{V}^{-1}(\bar{u}-\mu)\right] \qquad \text{(Eq 12)}$$

What is important to notice is the $\boldsymbol{V}^{-1}$. *Computing the probability requires the inverse of the covariance matrix.*

Now typically, a spectral estimate is going to use some number of points between 2^7 and 2^{15}, so with our table *h(t, n)*, the number of columns *N* could be as high as 32,768. The covariance matrix ***V***, since it's $N \times N$, is going to be enormous. Leaving aside for the moment the sheer number crunching this implies, one must suspect that there's some superfluous arithmetic in the computation.

Even by eye, the original waterfall tableau of spectral data shows a lot of redundancy. Many of the columns are very highly correlated. Some of them are even indistinguishable. The covariance matrix is going to be populated by values with only a minority contributing significant information. This gives us some confidence that the measurement vectors can be mapped onto a set of lower-dimensional vectors without losing anything important. The low-dimensional vectors are the feature vectors and they are points in feature space, just as the measurement vectors are points in the measurement space.

Among the statistical classifiers, the chief differences arise over how the mapping of measurement vectors onto feature vectors is accomplished and how the covariance matrix of all observations is partitioned into clusters. You will discover the full variety of approaches yourself, if you decide to experiment on your own using any of the existing software packages that support classification and clustering.

We should also notice that great promise has been shown by prior reduction of the observation space by a variety of techniques. One avenue is to apply "lossy" compression techniques like wavelet transforms or iterated fractal analyses to the measurement vectors before any further computations, as a way of eliminating known redundancy in advance of any statistical measurements. This is still a relatively unexplored area, one where an individual might easily make a fundamental contribution.

A Final Proviso

It was suggested before that the probability models described here are not the most powerful available. These models treat the observation vectors—either the measurement or feature vectors—as independent, meaning their order of appearance is generally not considered. You could shuffle them and the result would be the same, and it would be adequate for most signal classification tasks.

The accumulating sequence of likelihood computations leading to an assignment decision reflects nothing about the original ordering of the training data. Incorporating markovity in the models attains an extraordinary increase in power. Dynamic programming and Hidden Markov Modeling are two closely related approaches to this formidable task. It's a reasonable conjecture that the universal classifier will, in broad outline, resemble a network of Hidden Markov Models, whose organization and structure are acquired by the more freewheeling capabilities of machine learning. □□

D-STAR, Part 3: Implementation

We've seen the "whys" and "hows" of D-STAR. Let's look at the hardware and possible uses for the system.

By John Gibbs, KC7YXD

This article, the final part of the series, investigates the block diagram and performance of the prototype equipment to better understand the design issues of a D-STAR digital radio.

The hardware used in testing the D-STAR standard is shown in Fig 1 and the performance of the mobile unit is summarized in Table 1. Some of this hardware is available today and we expect several manufacturers will offer hardware soon.

Recall that the D-STAR standard has only recently finalized the selection of the modulation and codec. Prototype testing demonstrated that GMSK modulation and the AMBE 2020 codec gives the best combination of spectral efficiency and robust communications.

The IF and RF parts of the block diagram (see Fig 2) of the ID-1 shows a straightforward dual-conversion superheterodyne design that should look familiar to those experienced with analog rigs. However, several issues in a digital-radio IF are not clear from the block diagram.

IF Design Issues

The first issue with digital-radio IFs is that the group delay of the IF structure is critical. While analog radio designers can ignore phase linearity, group-delay variations need to be less than about 10% of the data period to avoid excessive BER due to intersymbol interference.

The second issue with digital radio IFs is that IF bandwidth must be wider than that of an equivalent analog design. It must be wider so that significant energy does not fall near the band edges of the filter because there the group delay is not constant.

Table 1
ID-1 Specifications Summary

Operating frequency	1.2 GHz Amateur Radio Band	
Operating Modes	FM (analog voice)	
(FDMA)	0.5GMSK (digital voice / data)	
Data Rate	4.8 kbps (voice) / 128 kbps (data)	
CODEC	AMBE	
Data Interface	IEEE802.3 (10Base-T)	
RF Power	10 W/1 W	
Receive Sensitivity	FM	–16 dBu
(typical)	4.8 kbps GMSK Voice	–10 dBu
	128 kbps GMSK Data	+ 2 dBu
Switching time	10 mS (digital mode)	
GMSK Modulation	Quadrature Modulator / FPGA (baseband)	

18225 69th Pl W
Lynnwood, WA 98037
kc7yxd@arrl.net

It often rises significantly and displays what are called "ears" (from their shape). This is particularly true in the receiver IF where transmitter and receiver relative-frequency tuning errors may cause the signal to be off center in the IF. Unfortunately, this increases the noise and interfering signals that pass through to the detector.

The quality of these IFs is measured by the sensitivity numbers in the specifications and in the "eye" diagrams in Fig 3. The well-open eye means that the receiver can easily distinguish between the plus and minus signal sent and therefore decode with very little BER. Fig 4 shows how the BER improves as S/N increases in the digital voice mode.

The final issue with digital radio IFs is the quality of the local oscillators. First, as implied above, the frequency reference must be accurate and temperature-stable if communication is to be established at UHF with a reasonably wide receiver IF. Second, the close-in phase noise of the local oscillators must be kept low, particularly if QPSK and other high-data-rate modulations are used. Excess oscillator noise can increase the BER just as effectively as actual channel impairments. One of the advantages of using GMSK is its relatively low sensitivity to these receiver problems, as shown in Fig 5.

Baseband Design Issues

The baseband hardware and modulators have far more obvious differences in this digital radio block diagram in Fig 2. For instance, on the transmitter side, the audio input is immediately converted to digital form, even if the radio is in the analog FM mode. This digital information is then signal-processed digitally and modulated onto the first IF. The modulation is accomplished by an I/Q modulator made with an FPGA. When teamed with DSP, an I/Q modulator is a very versatile component that can handle any form of modulation needed in the ID-1. It is even possible to produce narrowband-FM with the digitized voice. (The analog FM feature is desired for compatibility with existing analog radios.)

D-STAR Applications

D-STAR is very much a "blank slate," waiting for amateurs to write upon it. We can exploit its capabilities for a variety of old and new uses. Here are a few of the many suggestions we have heard from the Amateur Radio community as possible applications of D-STAR.

Mobile and Portable Internet Access

The application that springs almost instantly to everyone's mind is high-speed wireless Internet access. Part of the reason is that the Internet has become such an important communication and information tool in hams' lives today. Another reason is all the hype built up around third-generation (3G) cell phones and the DOCOMO system in Japan. Yet, with today's meltdown in telecommunication commerce, it could be years before a 3G phone system is deployed in the US. So, with the deployment of D-STAR, hams could once again have a leading technology that the rest of the population would envy and that might encourage more people to get their tickets.

In support of this vision of D-STAR as an Amateur Radio community growth agent, it is interesting to watch the reaction of inactive no-code hams. For a variety of reasons, they got their tickets, but never really got interested in the hobby. Often when they see a D-STAR demonstration, you can see their eyes light up and almost hear the gears turning in their head! Several have said that a system like D-STAR would get them active again.

Because this is Amateur Radio, there will be some restrictions on this vision of high-speed wireless Internet. The FCC does not allow encryption, so there is no guarantee of privacy. Anyone can look over your shoulder and read your e-mail.

Some hams bring up the issue of advertising and pornography. Control operators will be responsible for the content passing through their repeaters exactly as they are today. However, this does not seem to be a very difficult issue. Inexpensive software exists today that can filter out this offending material. Control operators can easily incorporate so-called "kiddy filter" software into the repeater's Internet interface. If the existing software does not quite fit our application, then resourceful hams will develop better software!

Then there is the issue of third-party traffic. Again, the control operator is responsible for ensuring that no illegal third-party traffic passes through his or her station.

Fig 1—Currently available hardware (counterclockwise, from upper right): RC-24 Control Head, ID-1 1.2 GHz transceiver, ID-RP1D 1.2 GHz data repeater, ID-RP1VS 1.2 GHz voice repeater, ID-RP1L 10 GHz backbone repeater and AH-1045/1080 parabolic antenna.

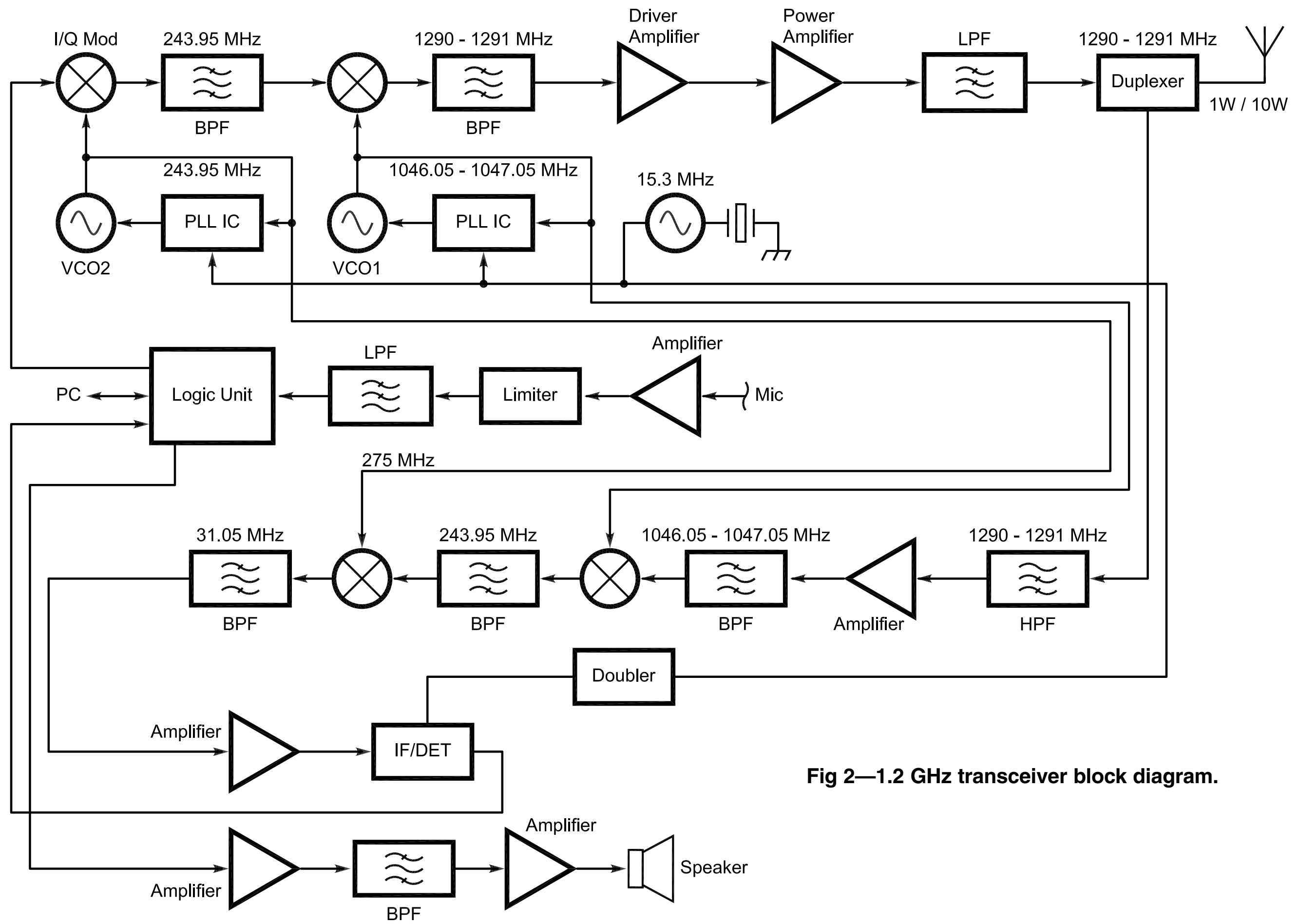


Fig 2—1.2 GHz transceiver block diagram.

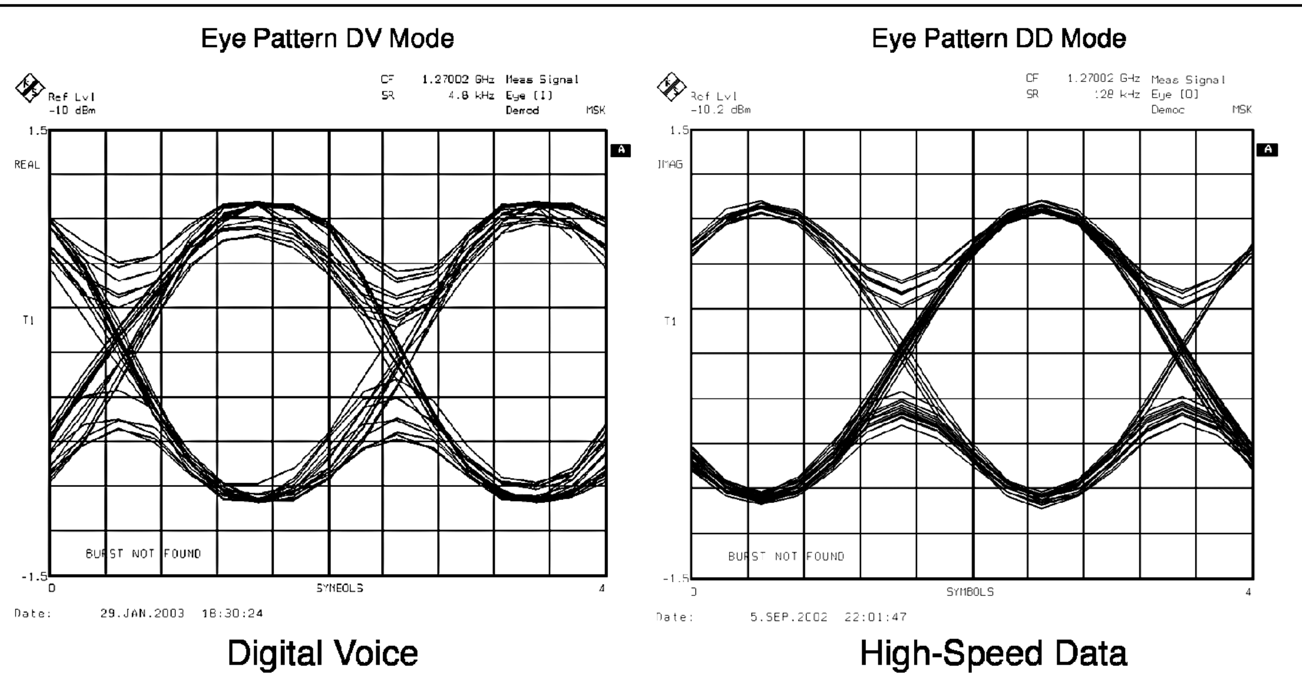


Fig 3—"Eye" patterns for digital voice (A) and high-speed data (B).

Combination of 802.11 and D-STAR

It would be surprising if manufacturers did not quickly develop a handheld D-STAR-compatible radio, but the high-speed data mode will necessarily have reduced range compared to a mobile rig with a good antenna and more power. What could you do if you wanted to connect a notebook computer to the Internet, but you are beyond the limited range of a handheld?

When hams have a range problem with handhelds on today's analog FM system, they sometimes cross-band repeat using their car's mobile radio. A similar solution could be implemented for high-speed data using D-STAR and a wireless LAN access point. A D-STAR mobile in your car could be cross-band (and cross-mode) connected to an access point installed in your car. Only an Ethernet cable is needed for this connection (no PC). If you already had a wireless LAN card in your notebook computer, you would be ready to go. Your notebook computer now has high-speed Internet access with the range of the high-power mobile.

Other High-Speed Data Applications

The Internet is so pervasive today that we sometimes forget that there are many other uses for high-speed data transmission. Here are two high-speed data applications that have arisen in D-STAR discussions.

Local Amateur Intranet: Rather than connecting to the Internet, a club-sponsored repeater could offer a wireless, wide-area Intranet. What might they put on the site? It certainly is a good place to make available the repeater system's operation guide and rules. To encourage D-STAR experimentation, it would be useful to have posting of hams' experiences with the system as well as freeware and shareware that they have found useful in D-STAR operation.

Visitors to the area could download information they need, even at 3 AM. Are you looking for a good Mexican restaurant, or do you need a quiet motel away from the highway? The Intranet could have suggestions from other hams on file, and you could download maps, driving directions and even pictures.

The possibilities multiply enormously if your notebook computer has GPS. Now D-STAR can guide you exactly to your destination with accurate maps and directions that better reflect the local driving conditions than those provided by major services on the Internet. Local hams could help you avoid traffic problems caused by temporary road closures and accidents.

Emergency Communications: Another D-STAR Intranet application is emergency communications. Even if the local D-STAR repeater were knocked out, temporary repeaters could quickly be assembled using two transceivers back-to-back. Training needs are minimized by using standard Internet browsers. When an operator comes onto the system, he can easily access stored files and bring himself up-to-date on the situation without distracting others.

Possible Add-Ons and Enhancements

We wrap up this discussion of the new D-STAR system with a treatment of the possible directions in which applications might evolve. D-STAR is not meant to be a turnkey communication system like the cell-phone system. Instead, it is an infrastructure that hams can use to meet current and future communication needs. Most importantly, it is a flexible, highly capable system that allows amateurs, them-

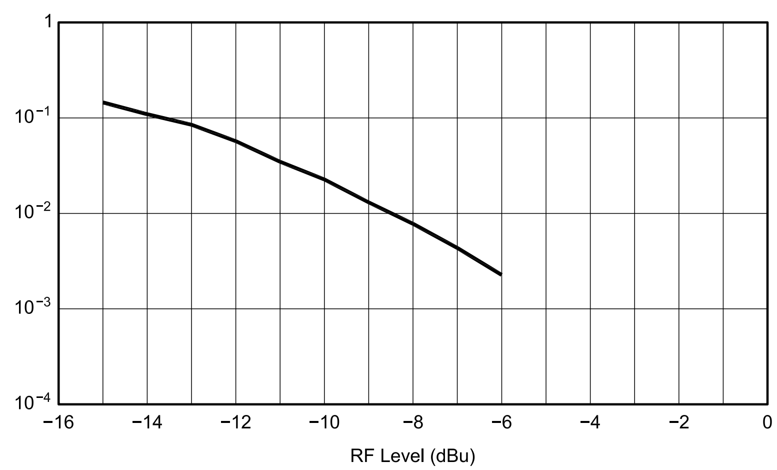


Fig 4—Bit-error rate versus RF level.

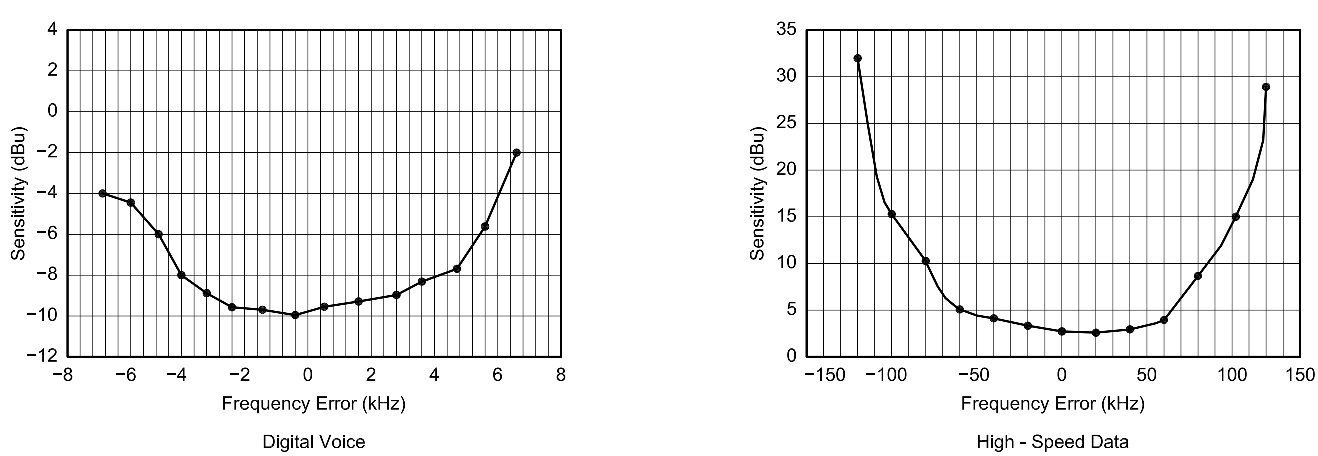


Fig 5—Frequency error versus sensitivity.

selves, to expand their service. Rather than depending on manufacturers to provide new features and applications, we expect the amateur community will develop add-ons to the system that will address the major goals of Amateur Radio including emergency communications, experimentation and just plain fun! Hams who have seen early demonstrations of the D-STAR system have generated the following ideas.

Power to the People!

We hams have our own opinions of how products should be designed and which features should be added. One of the great things about the D-STAR system is that for a large part, it is possible for us to try out our ideas and further the state of the art. Error correction is one area that is ripe for contribution by hams.

As data rates increase or as we push the range, decoding errors begin to be significant in any digital radio system. This is less of a problem for a properly designed digital voice system, because it is not significantly disturbed by BER levels that would render digital file transfers impossible. Yet any high-speed digital mode can use any help it can get.

Because of the importance of transferring data quickly and accurately, there has been a great deal of theoretical work done on coding and error correction. Tom McDermott, N5EG, gives a good introduction to the many coding techniques used in digital radio including Reed-Solomon, Golay and convolutional codes.[1] Newer codes called "turbo" codes have been developed that approach the theoretical limit on how fast information can be transmitted over a noisy band-limited channel.

However, these codes are only optimal if the interference is what we call AWGN (additive white Gaussian noise). This is true because the mathematics of AWGN is well understood. The bad news is that most of the impairments we find in real radio communications do not match this nice mathematical model. The good news is this is an opportunity for Amateur Radio to again advance the state of the art.

An interesting example of the possibilities of error correction is the ubiquitous CD player. A few years ago, I saw a demonstration of the power of the error correction used in CDs. The professor had drilled large holes in a CD and despite these obvious faults in the data stream, the music played perfectly without a click, pop or drop out! Perhaps some enterprising ham can discover the way to make just as dramatic improvement in radio.

Interleaving Spreads Bursts of Errors

Wireless communication channels with fades of the signal power are prone to errors occurring in bursts. Burst errors can cause problems by breaking error-control codes when the number of errors exceeds the maximum number of correctable errors for the specific code used. For short bursts, intraframe interleaving improves performance by spreading the burst of errors over several different code words. For example, if four code words, each containing 23 bits that can correct up to 3-bit errors, are used in a frame consecutively, then a burst error 4 bits long will break a single code. However, if the four code words were interleaved (that is, bit 1 codeword 1, bit 1 codeword 2, bit 1 codeword 3, bit 1 codeword 4, bit 2 codeword 1, and so forth), each code word would contain only one error, which could easily be corrected. Since intraframe interleaving only modifies the bit ordering within the current frame, no additional delay is generally needed for implementation. If additional delay can be tolerated, interframe (more than just the current frame) interleaving can be used to further increase the performance with longer burst errors.

Mixed Voice and Data

As we saw in the section on the D-STAR standard, the proposed digital voice protocol has the ability to transmit low-speed user data simultaneously with voice. The first-generation D-STAR transceivers minimally support this feature. However, as new radios are introduced, it is expected that hams will develop applications that exploit this capability. Notice that in the D-STAR system, this is referred to as low-speed data. Yet the data rate is actually about 2400 baud, faster than the old 1200 baud of amateur systems (and yes, slower than the 9600 baud used in higher-speed systems).

What could we do with this feature? How about the equivalent of the Internet's "instant messaging"? With instant messaging, messages could be added by the sender or even from a third party (where legal) and added by the repeater. Imagine that you are in the middle of a contact when

- A DX alert displays on your mobile for a country you need, or
- A printer attached to your transceiver prints out route instructions to your club's Field Day site, or
- Your spouse sends you the grocery list and reminds you that the lawn needs mowing—well, maybe that isn't such a good feature!

How about doing instant messaging one better and send instant pictures. The miniature cameras used recently in cellular phones are about 96×96 pixels; that is less than 10 kbits. So, a picture could be sent in less than 30 seconds simultaneously with a voice contact.

In a sense, this voice and data capability is like DSL: you can talk over the same channel while data are transmitted—although not at DSL speeds in this mode. The data you can send through this channel are limited only by your imagination. For instance, what do you think about mixing voice and next-generation APRS?

VoIP and D-STAR

VoIP voice communication is of course possible in the digital data mode because it does not matter what information is carried in the data. However, VoIP is not a very attractive method of communication via Amateur Radio today. It often suffers from poor voice quality scores due to the very long latency from the intensive signal processing and because the Internet does not give priority to voice packets.[2] These voice quality problems would certainly not be helped by the 128 kbps data rate of D-STAR.

Finally, VoIP on D-STAR is spectrally inefficient, requiring 130 kHz of bandwidth compared to less than 6 kHz for the highly compressed D-STAR digital voice mode. Still, for applications that require higher-speed data simultaneous with voice, inventive amateurs may find solutions to these problems.

Registration

The D-STAR proposal currently keeps a list of amateurs (call signs) who have accessed the system. So, if you want to call me, KC7YXD, you don't need to know the linking repeater. The system simply finds the repeater I last accessed and automatically routes your call to me. A logical extension of this capability is that if my radio is on the repeater frequency, the system can poll it and automatically register me onto that repeater.

This feature could be extended to keep a database at each repeater of each registered amateur's interests. How would our hobby change if you could call "CQ Collins radio collector" and automatically link to someone on the other side of the country or perhaps the other side of the world?

Roaming

Another feature that hams could add to the system is roaming. What if,

[1]Notes appear on page 47.

when driving through an area, the repeater could download into the radio memory the frequencies and call signs of nearby D-STAR repeaters? Then as I drive away from the repeater, the radio is all set to access the next repeater. Never again sit down with a repeater book and program the radio before the next trip!

Of course, if we were to add GPS capability and the D-STAR repeater database held the footprints, calls and frequencies of adjacent repeaters, the radio could *automatically* switch repeater sites as you drive through an area!

Trunking

"Trunking" is a land-mobile-radio term for a system that uses multiple repeaters to support many contacts at once. Most trunking systems use a "home channel" for calling, then the system assigns a clear repeater frequency to complete the contact. The radios then automatically go to the assigned frequency. The basic advantage of trunking is that the system can support many more users simultaneously than with individual systems. Effectively, it lets one listen to all repeaters in an area by only monitoring the home channel for a call. Since the D-STAR system sends call signs digitally, it is easy to envision a simple computer program that would monitor the home channel and alert me when I am being called.

Conclusion

Clearly, Amateur Radio is at a crossroads today. Technical and regulatory forces are pushing us out of our well-proven but inefficient ways. The possibilities that digital radio brings to our hobby are truly limited only by our imagination.

I hope this article has stirred your imagination and stimulated your interest in the possibilities of digital voice and high-speed data in Amateur Radio today. Perhaps you will be inspired to try the D-STAR system and maybe even develop applications or variations of the D-STAR system.

Recommended Reading

Visit **www.dvsinc.com** to read more about AMBE and to hear voice samples at various coding rates.

D.W. Griffin and J.S. Lim, "Multiband Excitation Vocoder," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol 36, No 8, August 1988, pp 1223-1235.

Notes

[1] *Wireless Digital Communications: Design and Theory,* Tom McCermott, N5EG, Tucson Amateur Packet Radio Corporation, 1996.

[2] The D-STAR digital voice mode addresses this problem by giving real-time data, such as voice, priority over repeater links.

At age 3, John exhibited early talents in electronics by "helping" his dad fix a TV. He plugged the speaker into a wall socket! Despite this traumatic start, he spent his youth building Heathkit and EICO equipment, repairing vacuum-tube radios and TVs and designing and building numerous homebrew projects including a Morse decoder high-school project built with resistor-transistor logic in the mid-1960s.

With BSE and MSEE degrees in control and communication theory, he has worked for Hewlett-Packard in the fields of spectrum and network analysis and frequency synthesis. He is currently the research department engineering manager at ICOM America, where his primary interests are digital communications and DSP. John has eight patents and is currently applying for four more.

An Extra class license holder, John usually is found on the HF bands, primarily operating PSK31. □□

The EMI Finder

Easily locate sources of electromagnetic interference with this sensitive, low cost, UHF receiving system.

By Richard Kiefer, KØDK

This article describes a UHF receiving system used to easily locate a very wide variety of sources of electromagnetic interference (EMI). The system consists of a sensitive narrow-bandwidth 318 MHz receiver, a directional handheld Yagi antenna and a headset (see Figs 1-3). Some of the typical broadband emitters of radio frequency interference (RFI) that you can pinpoint are power-line insulators, computers and industrial equipment.

The following are the principle technical features of the receiving system, which contribute to its effectiveness:

- High gain, directional five-element hand-held Yagi antenna for precise position location of noise emitters (see Fig 2).
- Simple construction.
- Surface acoustic wave RF filter (SAWF) for good front-end selectivity.
- High-gain, low-noise figure RF preamplifier (LNA) for high sensitivity.
- Regenerative detector stabilized with a surface acoustic wave resonator (SAWR) produces a very narrow detection bandwidth hence a low noise floor.
- High detection sensitivity, –136 dBm minimum discernable signal.
- Low-loss direct coupling technique between the LNA and regenerative detector.
- Low-voltage, low-current operation for long battery life.
- With only two active RF components, the circuit is simple and the cost low.
- All surface-mount devices and components for best RF performance, easy construction and low cost.
- Receiver design technique useful at other frequencies including the 70-cm amateur band.

4700 47th St
Boulder, CO 80301
k0dk@arrl.net

Find Broadband Sources of EMI

You can locate sources of electromagnetic interference by listening in the VHF/UHF range for their broad emission spectrum, which typically extends to 300 MHz and sometimes well beyond. Such spurious noise is usually amplitude modulated or contains both AM and FM.

For example, power lines, which are often a source of amplitude-modulated EMI put out noise into the UHF range from corona and sparking. These problems can be caused by defective hardware such as insulators and ungrounded transformers, which can act like spark gap transmitters radiating electromagnetic energy with a very broad spectral content. This broad spectrum usually contains enough energy at 318 MHz to be detected by a sensitive AM receiver such as the one described here. It is common practice in the power industry to search for defective power-line equipment by listening in the 300 MHz range with sensitive receivers and directional antennas. Commercial gear used to search for such EMI is available from companies like Radar Engineers (**www.radarengineers.com**) or Trilithic Incorporated (**www.trilithic.com**). Prices for commercial equipment range from several hundred to several thousand dollars.

Most digital electronic devices such as personal computers, games, DSL lines, VCRs and electronic instruments can also radiate a spectrum with components extending to above 300 MHz. Even if such equipment is compliant with the FCC Part 15 class B regulations for residential use, this receiver will usually find them if you are within 3 to 15 feet. Such equipment often have internal microprocessor and DSP clocks running in the 10 MHz to 100 MHz range, which create harmonics to beyond 300 MHz. Digital bus speeds are usually in this range as well and may vary in frequency and amplitude as software performs operations on peripherals and memory chips. All of this can produce a very rich spectrum at 200-300 MHz and above.

At my particular location, I can "hear" computers, displays, printers, a digital scope, a microwave oven controller, power lines, power substations, automobile ignition systems and garage-door opener superregenerative receivers. Around town, I can walk into most any building, sniff around and find most active electronic devices. All of these sources of RFI are broadband in nature. That is, their emissions extend over many megahertz and to very high frequencies. This receiving system, though, is a very narrow-band device, which listens to only a small slice of the total spectral output of any RFI emitter. The bandwidth of the five-element Yagi antenna is only a few percent of 318 MHz, and the receiver passband is only a few kilohertz wide.

A receiving frequency of 318 MHz is chosen for three main reasons. First, at this frequency the antenna is small and easily transported. It also has high gain and sharp directivity with only five elements. In addition, 318 MHz is a generally interference-free frequency in the USA. Although garage-door openers and automotive keyless entry systems operate at 315 MHz, no other radio services use 318 MHz. The harmonics of the lower frequency TV stations and communications radios are elsewhere.

Fig 1—The author searching for power-line noise.

Thirdly, there are off-the-shelf SAW filters and SAW resonators readily available for 318 MHz.

A High Sensitivity Receiver

To detect a very low-level slice of a broad-spectrum emitter you need a sensitive narrow-bandwidth receiver. The receiver described here uses a unique, but simple, RF circuit design to achieve sufficient sensitivity to locate noise sources. It has a minimum discernable signal sensitivity of about –136 dBm when using headphones, corresponding to a bandwidth of a few kilohertz.

Consider the following: The five-element Yagi antenna has a source impedance of 50 Ω at 318 MHz. So, as a noise generator its equivalent circuit is a voltage source in series with a 50 Ω resistor. If we measure the output noise of the antenna with a high-impedance measuring device that has a known 3-dB bandwidth, we will measure the following RMS voltage.

$$e_n = \sqrt{4kBTR} \qquad \text{(Eq 1)}$$

Where

e_n = The open circuit RMS noise output of the antenna.

k = Boltzmann's constant.

B = Bandwidth of the receiving or measuring device in hertz.

T = Temperature of the antenna noise resistor in kelvins.

R = Antenna resistance in series with the voltage source.

This equation determines the lowest possible noise floor, hence sensitivity, of the receiver assuming perfect noise figure of 0 dB (the noise figure of the receiver preamplifier is actually about 2 dB). If we substitute a signal generator with a 50-Ω output

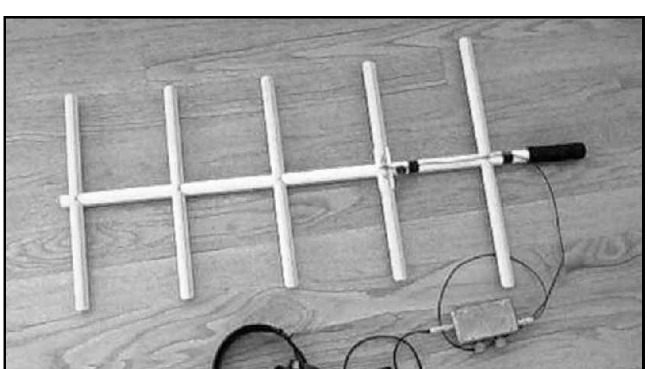


Fig 2—The complete receiving system: 318 MHz Yagi, receiver and headset.

Fig 3—The 318 MHz receiver. The controls are for regeneration and volume.

impedance for the antenna, we can determine the approximate bandwidth of the receiver. If we insert a signal level that is equal to the noise level then the receiver bandwidth can be calculated from:

$$e_n = \sqrt{4 \times 1.38 \times 10^{-23} \times B \times 293 \times 50} \quad \text{(Eq 2)}$$

Where
e_n = the generator RMS voltage.
B = equivalent bandwidth of the receiver.

For this receiver a 100% AM modulated generator at –133 dBm (50.2 nV) produces a signal that is about equal to the noise. So, the 3 dB bandwidth of the receiver is about 3.1 kHz, and the minimum discernable to the human ear signal level is about 3 dB lower at –136 dBm. Not too bad for a circuit with only two active RF parts!

Receiver Circuit Description

The receiver circuit works as follows. Refer to the schematic diagram in Fig 8. A signal or noise in a narrow band at 318 MHz is picked up by the Yagi antenna, which has a source impedance of 50 Ω. The antenna output voltage is applied to the input of a SAW bandpass filter centered at 318 MHz. The filter rejects adjacent-channel interfering signals, which could overload the preamplifier causing spurious responses—not the noise you are looking for. Such interfering signals could be TV transmitters below 300 MHz, transmitters in the 70-cm Amateur Radio band (420-450 MHz) or other communications radios. L1, L2, C28 and C27 provide impedance matching for the input and output of the SAW filter. L5 is an RF choke for front-end ESD protection. The output of the SAW is then amplified by a Motorola preamplifier chip, the MBC13916. This is a cascode preamplifier with 22 dB of gain and a noise figure of 2 dB. It is a silicon-germanium part that contains a temperature-compensated internal bias circuit.

The high-impedance output of the preamplifier is summed by superposition in the collector of the regenerative detector, Q1. This is a very straightforward current-source summing technique that eliminates a coupling capacitor or other parts that might load the collector of the regenerative detector. The circuit of Q1 is the regenerative detector, which is a common-base oscillator stabilized by a 318 MHz SAW resonator (SAWR). The SAWR acts like a 318 MHz fundamental-mode crystal, with a similar equivalent circuit. It serves to provide excellent short-term oscillator

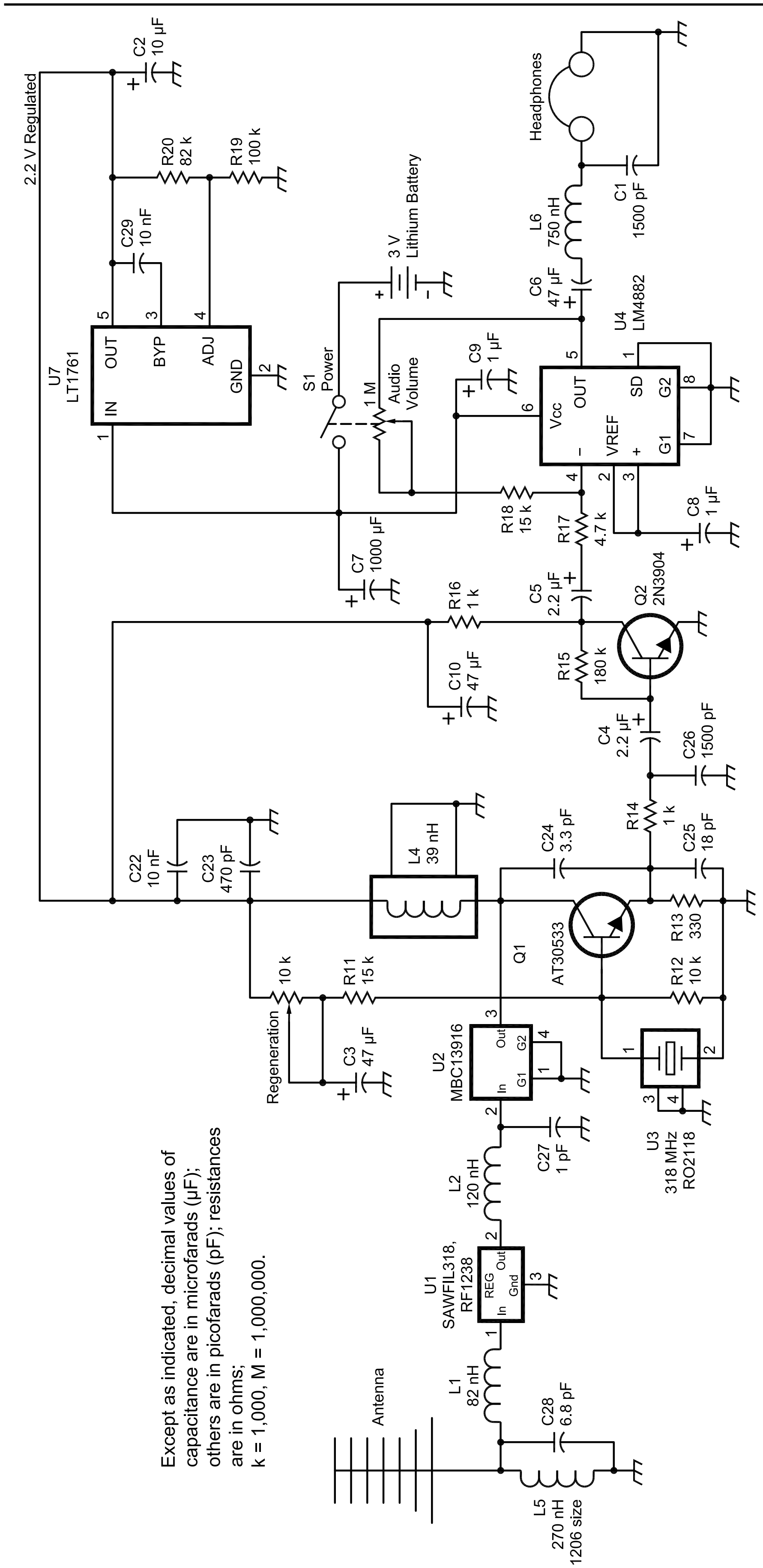


Fig 4—The schematic diagram of the EMI Sniffer.

stability because of its high Q. The unloaded Q of the SAWR is specified at 16,900 by the manufacturer, RF Monolithics. This is good although not as high as a crystal, which is typically greater than 20,000. The circuit loaded Q of the SAWR will be less than 16,900 but still high enough to produce a very narrow receive bandwidth when used as the frequency determining element of a regenerative detector.

The frequency tolerance of the SAWR is ±100 kHz, and the long-term and temperature stability is on the order of several tens of kilohertz. Here, only the short-term stability of the SAWR, and hence the regenerative detector, is important. For this application we can listen to any frequency near 318 MHz as long as the short-term stability is good enough to produce a low noise floor. This assumes that the noise we listen to is uniformly distributed over a 200 kHz bandwidth centered on 318 MHz, a pretty good assumption for most EMI emitters. The short-term stability of the detector is determined by its loaded Q, which is something less than 16,900 multiplied by the regenerative effect of the detector circuit. If the effective bandwidth of the receiver is 3.1 kHz, as discussed above, this corresponds to an effective Q at 318 MHz of 102,600.

The 18 kΩ potentiometer fixes the onset of oscillations of the regenerative detector, hence the most sensitive detection point, and is adjusted by the operator with a knob (see Fig 4). The output of the regenerative detector is taken at the emitter, low-pass filtered and preamplified by the audio stage, Q2. The output of the preamplifier is then passed to the National Semiconductor power amplifier chip, the LM4992 and boosted to a level sufficient to drive a headset or small speaker. The audio volume is adjusted with the 1 MΩ potentiometer. The whole circuit is powered by a 3 V CR123A lithium battery. This battery is typically used in film cameras, and it is readily available. The 3 V output of the battery is regulated down to 2.2 V by the Linear Technology LT174 low-dropout regulator. The receiver circuit is designed to operate from 2.2 V because this is about the end of life for the battery.

The Five-Element Tape-Measure Yagi

The ability to precisely locate sources of EMI is enhanced by the five-element hand-held Yagi antenna used with the receiver. The antenna improves the sensitivity and utility of the system because of its high gain and directionality. The beamwidth is quite narrow as you can see from the computer simulation of the pattern in Fig 6. A narrow pattern makes it easy to peak up the noise from an emitter as you swing the antenna back and forth, up and down.

Changing the polarity from horizontal to vertical also sometimes improves the strength of the noise. Usually you can pinpoint a source of noise within inches if you are five feet away. From ground level, you can locate defective insulators on power poles.

The Yagi is computer optimized for the best pattern and impedance match using the computer program Yagi Optimizer (*YO*). The front-to-back ratio of 25.2 dB is adequate for good direction finding. [Fig 6 was generated with *YW*, a similar program packaged with *The ARRL Antenna Book*. The

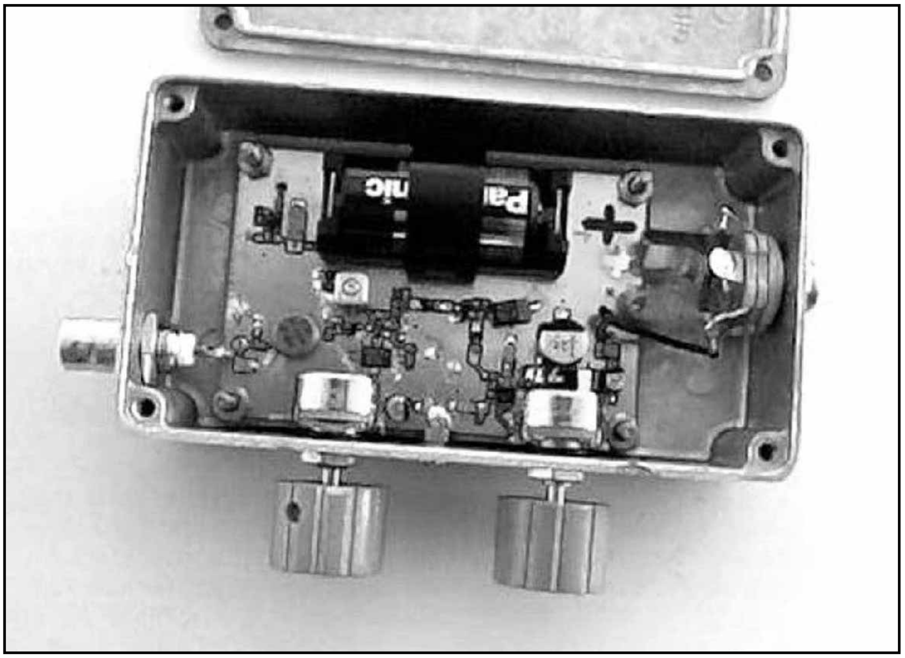


Fig 5—The receiver printed-circuit board showing the regeneration and volume controls.

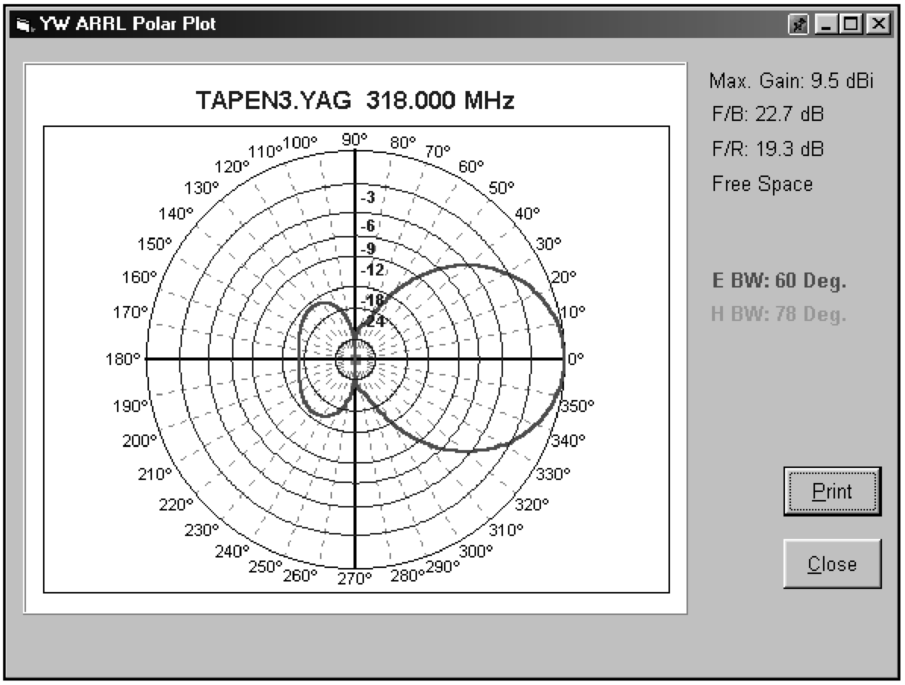


Fig 6—Computer modeled pattern for the antenna.

performance varies slightly from that calculated by *YO*.—*Ed*] The 42 Ω feedpoint impedance (see Fig 7) presents a pretty good match to 50 Ω coaxial cable, such that the mismatch loss is very low. No matching network is used, just a single bead balun. The one-inch-wide spring-steel tape measure material is simulated in *YO* with a 0.15 inch diameter round aluminum element. This equivalent element diameter is determined by measuring the resonant frequency of a dipole made from the one-inch-wide tape-measure material. Then the actual length and resonant frequency are used as fixed parameters in the antenna modeling program *EZNEC* to find the equivalent diameter of a round aluminum element. In other words, the 1-inch-wide tape-measure material acts electrically like 0.15-inch diameter aluminum tubing at 318 MHz. So the elements modeled in *YO* are 0.15 inches in diameter and the lengths and spacings of the elements as calculated by *YO* are as shown in Table 1.

The boom of the antenna is made of CPVC, the hot-water version of PVC. It is stiffer than regular PVC and machines very easily. The elements are very rugged because of their flexibility, being made from steel tape-measure material. The driven-element mount is machined from Delrin using a 45° dovetail cutter as shown in Fig 8A. The cuts in the boom that hold the elements are also made with a 45° dovetail cutter. The elements are each fastened to the boom with a $\#8\text{-}32 \times 1^1/_2$ bolt and locking nut. If an element breaks, it is easily replaced in the field. The same antenna construction can be used for fox hunting on the 2 meter or 70 centimeter bands. Because of the flexible elements you can easily travel in close quarters and through bushes without difficulty. The antenna is very lightweight and can be carried for extended periods without fatigue. A foam bicycle-

Table 1

Yagi Antenna Dimensions in Inches

Element	*Length*	*Boom Position*
Reflector	19.0	0
Driven Element	17.2	7.5
Director #1	15.6	15.0
Director #2	15.4	22.0
Director #3	15.4	29.0

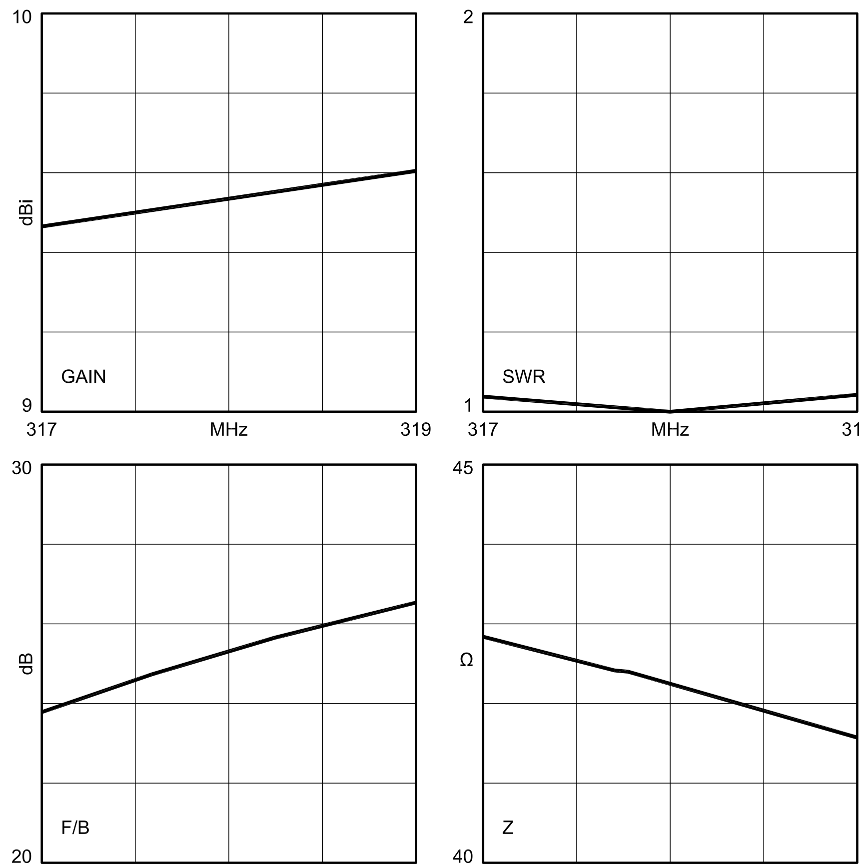
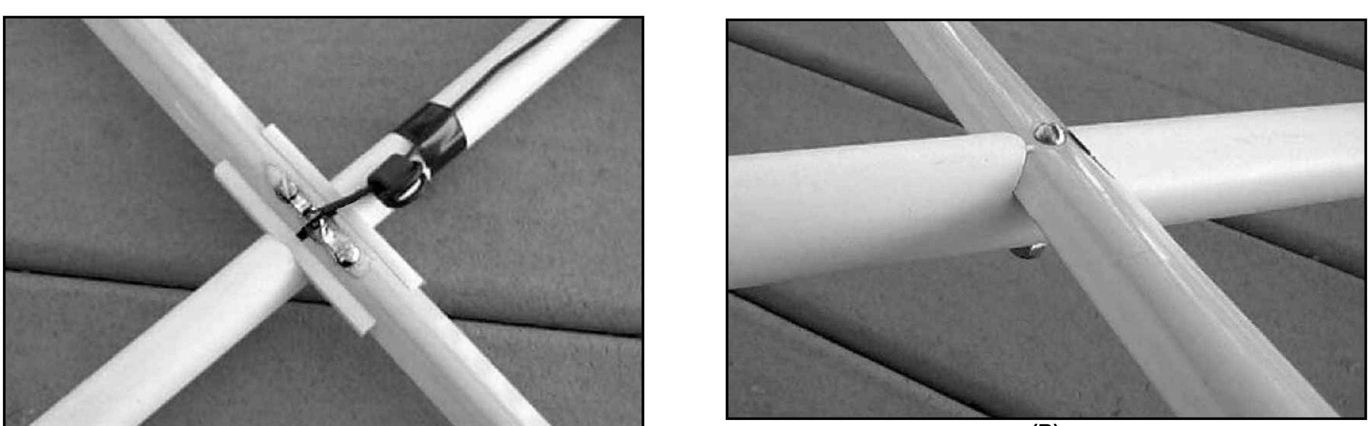


Fig 7—Computer modeled Yagi performance.



(A) (B)

Fig 8—At A, the feedpoint details. The mount is made from a $1^1/_4 \times 3 \times {}^1/_4$-inch-thick piece of Delrin plastic. It is machined with a 45° dovetail bit to accept the halves of the driven element, then drilled and tapped for three #8-32 bolts. The bolt positions are not critical. The two outer bolts secure the element halves to the Delrin mount, and the center bolt (fed through a hole in the boom) secures the mount within a notch cut into the CPVC boom. At B, milled dovetail cuts in the boom secure the tape-measure elements.

handlebar grip is used for the handle and it is quite comfortable.

Other Applications

The UHF regenerative receiver technique described here may be used in other applications as well. For example, you could build a simple AM or FM system to communicate voice or data over short distances in the 70 cm band. The receiver could be just as described, but use a commonly available SAWR at 433.92 MHz. The same resonator would also be used to control the transmit frequency. Since SAW resonators can be pulled about ±50 kHz in frequency, both the transmitter and receiver can be tuned to the same frequency. The circuits also operate at very low current levels for long battery life. It might also be possible to implement a microprocessor to automatically adjust the level of regeneration at the detector to maintain best sensitivity.

SAW devices are commonly used by the millions in automotive keyless-entry and garage-door opener products. A typical automotive keyless-entry system would use a SAWR stabilized transmitter in the key fob and an LC stabilized superregenerative receiver inside the instrument panel of the car. The transmitters are usually amplitude keyed with digital data at about 1 kbps. At least one keyless entry system that I know of is frequency modulated with a deviation of about ±30 kHz. You can pull a SAWR this far with a Varactor diode. You can also use other types of antennas to sniff for EMI on printed-circuit boards and enclosures. These could be loops, shielded loops and simple capacitively coupled tips. Or, you can make a direct connection to a circuit with a piece of coax. When attaching antennas and coax, observe static-protection procedures to not "static zap" or otherwise destroy the input SAWF. An input inductor to ground is provided, but this component will fail open if you apply more than a few hundred milliamperes to the antenna connector.

References

[1] C. Kitchin, N1TEV, and M. Murphy, WB2UID, "An Ultra Simple W1AW Receiver," *QST*, May 1997, pp 34-35.

[2] C. Kitchin, N1TEV, "New Super-Regenerative Circuits for Amateur VHF and UHF Experimentation," *QEX*, Sep/Oct 2000, pp 18-32.

[3] C. Cullian, KØRF, and T. J. Jasiewicz, "Finding That Elusive Power-Line Noise," *CQ*, Jan 1992, pp 28-31.

[4] J. Harris, Trilithic Inc, "Locating Power Line RF Interference," *RF Design*, Feb 1992, pp 42-51.

[5] RF Monolithics at **www.rfm.com** for SAWR and SAWF data sheets and application notes.

Bibliography

1. H. Ataka, "On Super Regeneration of an Ultra Short Wave Receiver," *Proceedings of the Institute of Radio Engineers* (IRE), Vol 23, No. 8, Aug 1935, pp 841-844.
2. E. Armstrong, "Some Recent Developments of Regenerative Circuits," *Proceedings of the Institute of Radio Engineers* (IRE), Vol 10, No. 4, Aug 1922, pp 244-260.

Richard Kiefer has worked as an electronics engineer since 1970 designing analog and radio frequency circuits and systems. He has also written software and firmware in several languages to control various circuits and hardware. He has worked for several electronic product development companies including IBM, Hewlett-Packard, Martin-Marietta and Armco-Autometrics. For the past 22 years, Richard has been the principal of Kiefer Electronic Development, an electronic product development consultancy specializing in radio frequency design. Richard holds BSEE and MSEE degrees, with academic emphasis in the subjects of RF, analog and linear systems analysis. He holds five US patents. Richard has been a licensed ham since 1959 and likes to work DX on 10-20 meters SSB with a Yagi stack on a 100-foot rotating tower. He is also working DX mobile through his EchoLink connected VHF repeater at Boulder, Colorado. □□

RF

By Zack Lau, W1VT

BUILDING MICROWAVE QUADS AND YAGIS

The lower microwave bands, 33 cm to 9 cm, represent a transition region where parasitic arrays such as quads and Yagis may be a useful alternative to the dish and horn antennas that are ubiquitous at higher frequencies. Their primary advantage is reduced wind loading. A bulky four-foot-diameter dish is unlikely to have as much gain as a 12-foot long loop Yagi.[1] Four feet is less than four wavelengths on 33 cm. Such a small dish is difficult to feed—a typical efficiency might be around 30%, compared to the 55% typically assumed. This is 2.6 dB less gain than one might assume from conventional charts and calculations. A well designed conventional Yagi can be even smaller—decades of computer optimization has resulted in designs that can generate 18-dBi gain with just an 8-foot boom.

Horn and dish antennas are more useful for space receiving applications, where you seek to maximize the gain to noise-temperature ratio. It isn't so easy to obtain high gain and clean pattern with a large parasitic array. Horns and dishs can be built with exceptionally clean patterns. Noise temperature can be a problem—Chip Angle told me he could hear EME signals off his newly constructed loop Yagi array. As the array got rusty, however, it got noisy, degrading his ability to hear extremely weak signals off the moon. It still works fine for terrestrial work.

For portable work, I've used a conventional Yagi for 903 MHz, an 11-foot-boom loop Yagi for 1296 MHz, a 5-foot boom loop Yagi for 2304 MHz and a 2-foot dish for 3456 MHz. I tried a long loop Yagi for 903 MHz, but the array was just too big to be manageable in high winds. The 2304 loop had decibels less gain than a 2-foot dish but much less wind load. At 3456 MHz, the dish has about a 4 dB advantage in gain. The dish is about seven wavelengths wide on 9 cm—the dish is getting big enough for practical feeds to be efficient.[2]

Home stations can generally put up bigger arrays for more gain. Typically, one can increase the gain by 2.5 dB by either doubling the boom length or array size. Home stations can benefit by using long 12-foot-boom Yagis for 2304 MHz, instead of a short 5 foot Yagi. At 1296 MHz, it becomes impractical to double the length of a long Yagi—it is more common to use a pair or quad of Yagis for more gain. Arrays of Yagis can be frustrating—a small non-obvious mistake can result in a severe loss of gain. For instance, flipping over Yagi will also invert the phase—a 180° phase shift is more appropriate for signal cancellation than addition!

Loop Yagis

Loopquad is the name preferred by G3JVL, who designed this antenna in 1974. However, in the USA, the name loop Yagi is almost universal, although it is confusing—how is a Yagi made out of loops different from a quad? It is a very distinctive and clever antenna. The loops are thin strips of metal bent into a circle and attached with screws to a metal boom. The driven element is fed with semi-rigid coax that passes a through the center of a ¼-20 brass screw holding the loop in place. This simple feed arrangement may also form a balun, reducing common mode currents on the feedline.

[1]Notes appear on page 56.

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

The original used a circular driven element, which results in a shield length around 1/3 wavelength. I've seen other designs with the element squashed down, so the shield length was closer to 1/4 wavelength. A 1/4 wavelength coaxial shield is ideal for balun operation—it inverts the low impedance present at the boom connection to high impedance at the feedpoint connection. Thus, unwanted current on the outside of the coax shield is minimized.

The most serious difficulty with the loop Yagi is the difficulty of modeling it accurately on a computer. Flat strips of metal aren't as easily modeled as round wires. Simple models are required—microwave loop Yagis have dozens of elements. A complicated model suitable for a three-element quad may not be practical with a 30-element loop Yagi. The computing time and resources rise exponentially with complexity—ten times the number of elements is likely to raise the time and resources required by a factor of 100, not ten. It may be even more, if the problem chokes the system. It could be less if the software were custom designed for this particular problem.

Fortunately, a few hams have optimized designs on test ranges, so there are several good designs available. I've had good luck with Chip Angle N6CA's 1296 loop Yagi.[3] Dave Olean, K1WHS, has designed high-gain 2304 and 3456 versions.[4, 5]

M. H. Walter, G3JVL, did some studies to allow scaling his original design to slightly different materials. I'm sure this helped the popularity of his design—it allowed both Americans and Europeans to adapt his designs to readily available materials. I'm sure it is just as frustrating for Europeans to deal with US customary sizes as it is for US hams to deal with metric sizes.

The companion software to the ARRL *UHF/Microwave Experimenter's Handbook* includes an old *Basic* program for adapting M. H. Walter's 27- and 38-element designs to different materials—you can change the element thickness, width and boom diameter. This is quite useful for translating designs between US customary and metric sizes—assuming you can deal with mid-1980s vintage software.[6]

Construction of loop Yagis may be difficult. The ideal tool for cutting the elements accurately is a sheet metal shear—something not found in the typical home shop. By cutting a sheet of aluminum into the proper width and then slicing it into strips, many elements of nearly identical length can be obtained. This may be a reason why many designs have just a few element sizes, compared with conventional Yagis that have no two elements of the same size.

The other difficulty is drilling the brass driven-element mounting screw. Brass is difficult to drill with normal drill bits—it tends to grab—sometimes pulling the work out of whatever is holding it. A zero-rake drill—one with a cutting edge perpendicular to the work, actually works much better on brass. A lathe is the preferred tool for drilling the hole—there isn't much margin for error. Drilling a 0.144-inch hole in 1/4-20 tapped brass leaves a wall thickness of just 15 mils. The drill must be centered accurately and the axis of the bolt must be aligned with the drill. A tolerance of 1° isn't good enough for two-inch bolts. You can use a little trigonometry to figure out the tolerance:

$$tolerance \quad \tan^{-1}\left(\frac{wall}{length}\right) \qquad \text{(Eq 1)}$$

where *wall* is the wall thickness and *length* is the length of the bolt. You may wish to narrow the angle even further, to account for the tolerance in centering the drill bit. A spotting drill, a thick stubby drill bit designed for starting holes, will help.

Yagis

I prefer to use conventional Yagis on 903 MHz—though I've not published any designs. A repeatable matching network has eluded my efforts at this frequency. I find a bit of tuning is necessary to get an acceptable SWR or return loss. I'm sure that proper fixtures or precision machining would help a great deal, but such a project won't be of interest to most hams.

In terms of wind load per element, rod elements are superior to loop elements. While a two-element quad has more gain than a two-element Yagi, the advantage disappears entirely as the boom length is made longer. In fact, long Yagis often have an extra decibel of gain, compared to long loop Yagis of the same length. This isn't surprising—Yagis have been optimized for decades using sophisticated computer techniques, while only a few dedicated experimenters have optimized loop Yagis on test ranges. At higher frequencies, Yagis become more difficult to fabricate, as the construction tolerances require more precise machining. Keith, KØKE, took 16+ hours to make a 3456 MHz Yagi, carefully filing the elements to ±1 mil accuracy.[7]

There are some pitfalls in modeling Yagis. Long Yagis modeled in *Mininec* typically suffer from a shift in frequency—to make them work properly you need to trim all the parasitic elements so that the actual elements are shorter than predicted in the computer model. This can be avoided by modeling them in *NEC-2*. *NEC*-2 isn't as useful at HF where tapered elements are common, but this limitation generally isn't a problem at microwaves. *NEC-2* is in the public domain, and can be downloaded from the Web. However, the user interface is archaic, even by 1980 standards, so buying a modernized version makes sense for most hams. Brian Beezley, K6STI, produced a proprietary version of *Mininec* that claims to eliminate the frequency shift, but he no longer advertises antenna modeling programs for sale.

Modeling falls short when it comes to typical matching networks, which are large enough to interact with the fields generated by the antenna. *NEC/Mininec* programs aren't adequate for modeling this situation. More advanced programs that use three-dimensional finite-element analysis may be up to the task, but these costly programs still aren't feasible for most amateurs. Thus, empirical design, or cut-and-try, is still the method of choice for accurately determining matching-network dimensions. It can be quite frustrating to come up with a good matching network—Avery Fine, KA3NTX, published a design with a 1.7 to 1 SWR, not having enough time to devise a better network.[8]

There are techniques to ease the difficulty of designing a matching network. The most obvious is to sacrifice a little gain and design the antenna for a direct 50-Ω feed. This avoids the need for a bulky matching network. If gain cannot be sacrificed, a slightly more difficult approach is to design the antenna for a direct 25-Ω feed and use a 1/4 wavelength of 35-Ω coax as the matching network. A simple balun that works well with such simple feeds is a metal boom that shields the coaxial feedline from the rest of the antenna—this works well with short, rear-mounted antennas. The coax can run out the back of the boom.

Mechanically, it is usually desirable to mount Yagi elements directly through an aluminum tube that acts as the boom. This is more reliable than mounting elements above the boom on plastic insulators—it is typical for insulators to degrade over time, resulting in an antenna that sheds elements whenever the winds get gusty. The shielding effect of the boom requires that the elements be made slightly longer. Thus, for a given element diameter and boom size, there is a certain length that needs to be added, known as the boom correction. The boom correction, which is almost negligible at 2 meters, becomes quite significant at microwaves. Fortunately, Guy Fletcher, VK2KU, has published a *QEX* article describing boom and element corrections for 1296 MHz.[9]

Especially for shorter Yagis, I prefer to just scale low-frequency designs, using *SCALE.EXE*. It uses the equations developed by James Lawson, W2PV, to recalculate equivalent elements for Yagi antennas. It can handle element tapers, in case you want to scale your favorite HF antenna. HF antennas typically use elements constructed out of telescoping aluminum tubing. *SCALE* is a DOS utility that comes with the *ARRL Antenna Book* software—not only does it allow you to change the frequency, but allows you to select more practical element sizes. For instance, you may want to make your antennas out of #8 aluminum ground wire or 3/16-diameter 6061-T6 aluminum rod. It is a good idea to check the scaled design with antenna modeling software, just to make sure that nothing has gone amiss. The best indicator is the level of sidelobes—it should be about the same as the original antenna. If the sidelobes are too large, you may need to make the elements smaller, by scaling them to a higher frequency. Conversely, you can scale the Yagi to a slightly lower frequency if the sidelobes are too small. *The UHF/Microwave Experimenter's Handbook* is an excellent reference for those wanting to design microwave Yagis, with articles by Steve, K1FO, and Gunter, DL6WU.

Testing

Measuring SWR is most easily done with a low power RF source and a microwave return loss bridge. Paul Wade, W1GHZ, published a simple design in *QEX*.[10] It features excellent directivity, over 30 dB from 10 MHz to 2.304 GHz. It is important to have good directivity—it is a measure of how low an SWR your meter can measure. According to Steve Polishen, K1FO, Bird wattmeter slugs typically have just 23 dB of directivity at 432 MHz.[11] The 70-cm band falls at the edge of the band limits for the slugs, rather than at the center, where one is more likely to get the typical 30 dB of directivity. A disadvantage of both is the use of a simple diode detector.

Paul used 1 kHz amplitude modulation to obtain some interference resistance—but you still need a clean RF source, free of strong spurious signals and harmonics. A low-power transmitter, legal for transmitting purposes, may not be adequate for serious measurements with a broadband detector.

One solution is to use a selective detector, such as a receiver or spectrum analyzer with a directional coupler. I've had good luck with surplus Narda directional couplers purchased at hamfests. They are easily tested with a good 50 Ω load and a reference short. Surplus dummy loads are easy to test with a dc-ohmmeter—if they are abused they will show a resistance significantly different from 50 Ω. Open circuited transmission lines can be rather tricky at microwaves—the center pin of a coaxial cable may look like an antenna rather than an open circuit.

The best way of getting a useful gain measurement is to take the antenna to a VHF or microwave conference that offers antenna gain measurements, such as the Eastern VHF/UHF Conference in New England and the Central States VHF conference in the Midwest. If it doesn't work, you often get helpful advice on what might have gone wrong. It is possible to set up your own range, but it takes a lot of careful work to get accurate measurements. An excellent tutorial on measuring antennas written by Dick Turrin, W2IMU, appears in *The ARRL Antenna Book*. It was originally published in the Nov 1974 *QST* as "Antenna Performance Measurements"—it is just as useful today as it was back then. Paul Wade, W1GHZ, expands upon his work in *The ARRL UHF/Microwave Projects Manual*, volume 2.

Notes

1. D. Hilliard, WØPW, "902-MHz Loop Yagi Antenna," *QST*, Nov 1985, pp 30-32.
2. Z. Lau, W1VT, "An Inexpensive 3456-MHz Dish Feed," *QEX*, March 1998, pp 53-55.
3. C. Angle, N6CA, "Loop Yagis for 1296 MHz," *ARRL Antenna Book*, 19th ed, pp 18-39 to 18-41.
4. D. Olean, K1WHS, "The Blow Torch 2304 and 3456 Loop Yagi," *Proceedings of the 25th Eastern VHF/UHF Conference*," pp 341-345.
5. D. Olean, K1WHS, "A Long, Long Loop Yagi for 3456 MHz," *Proceedings of the Microwave Update 2000*, pp 239-246.
6. K. Erickson, KØKE, "A 'First' 3456 MHz Long Yagi?," *Proceedings of the Microwave Update '87*.
7. A. Fine, KA3NTX, "A Low-Cost High-Performance 23 cm Yagi," *Proceedings of the Microwave Update*, 1989, pp 168-170.
8. G. Fletcher, VK2KU, "Effects of Boom and Element Diameters on Yagi Element Lengths at 144, 432 and 1296," Jan 2000 *QEX*, p 16.
9. P. Wade, W1GHZ, "A UHF+ SWR Bridge," *QEX*, Feb 1995, pp 3-5.
10. Z. Lau, W1VT, "10 GHz SMA to WR-90 Transition," *QEX*, Nov 1995, pp 24-26.
11. S. Powlishen, "On 432 #40 Common Sense Talk About SWR Measurement Accuracy," *Proceedings of the 27th Eastern VHF/UHF Conference*, pp 36-39.

2003 Index

Issues 216 through 221
(All dates are 2003 unless otherwise indicated.)

Upcoming Conferences

□□

United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title	2. Publication Number	3. Filing Date
QEX	0 8 8 6 _ 8 0 9 3	September 23, 2003
4. Issue Frequency	5. Number of Issues Published Annually	6. Annual Subscription Price
Bimonthly in: Jan, Mar, May, Jul, Sep, Nov	6	$24.00

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*
225 Main Street, Newington, Hartford County, CT 06111-1494
Contact Person: Katherine Capodicasa
Telephone: (860) 594-0257

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
225 Main Street, Newington, CT 06111-1494

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*
Publisher *(Name and complete mailing address)*
Mark J. Wilson, 225 Main Street, Newington, CT 06111-1494
Editor *(Name and complete mailing address)*
Doug Smith, 225 Main Street, Newington, CT 06111-1494
Managing Editor *(Name and complete mailing address)*
Robert Schetgen, 225 Main Street, Newington, CT 06111-1494

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
American Radio Relay League, Inc.	225 Main Street, Newington, CT 06111-1494

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → XX None

Full Name	Complete Mailing Address

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
XX Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526,** October 1999 *(See Instructions on Reverse)*

13. Publication Title: QEX
14. Issue Date for Circulation Data Below: Sep/Oct 02 - Jul/Aug 03 Sep/Oct 03

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*		6,483	6,400
b. Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. *(Include advertiser's proof and exchange copies)*	5,188	5,178
	(2) Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*	0	0
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	231	234
	(4) Other Classes Mailed Through the USPS	524	420
c. Total Paid and/or Requested Circulation *[Sum of 15b. (1), (2),(3),and (4)]* ▶		5,943	5,832
d. Free Distribution by Mail *(Samples, compliment ary, and other free)*	(1) Outside-County as Stated on Form 3541	81	84
	(2) In-County as Stated on Form 3541	0	0
	(3) Other Classes Mailed Through the USPS	71	31
e. Free Distribution Outside the Mail *(Carriers or other means)*		145	230
f. Total Free Distribution *(Sum of 15d. and 15e.)* ▶		297	345
g. Total Distribution *(Sum of 15c. and 15f)* ▶		6,240	6,177
h. Copies not Distributed		243	223
i. Total *(Sum of 15g. and h.)* ▶		6,483	6,400
j. Percent Paid and/or Requested Circulation *(15c. divided by 15g. times 100)*		95.24%	94.41%

16. Publication of Statement of Ownership
XX Publication required. Will be printed in the Nov/Dec 2003 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: [illegible], Publisher
Date: September 23, 2003

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form **3526,** October 1999 *(Reverse)*

Letters to the Editor

Efficient Energy Transfer Between Capacitors (Re "Energy Conversion in Capacitors," Jul/Aug 2003)

First is an explanation of where half of the energy goes when you connect a charged capacitor to an uncharged capacitor of equal capacitance. The circuit used for analysis is shown in Fig 1. The voltage source for charging C1 is shown as a battery, but a dc power supply could be used instead. A series resistor, R1, is included to limit the charging current. Capacitors C1 and C2 are assumed to have equal capacitances. Other values will be discussed later. The two leads connecting the two capacitors, through a switch or relay, S2, will have enough inductance to resonate the capacitors at some frequency. For example, 0.5 µH and two series connected 8 µF capacitors resonate at 112 kHz. The reactance of the inductance is $j0.35\ \Omega$. The effective loss resistance at this frequency, including dielectric losses in the capacitors, would be very much less than 0.35 Ω.

For Case A, let us introduce a very large resistance, R2, such as 10 kΩ, or more, making the inductive reactance negligible. Now, when S2 is closed, C1 will discharge through R2 into C2. Fig 2 illustrates how the voltages on the two capacitors vary over a period of three time constants, 3T. The voltage across R2 at any instant is V_{C1}–V_{C2}. At any instant, the power being lost in the resistor is $(V_{C1}–V_{C2})^2$/R2. Integrating the dissipation over several T periods shows that the energy lost in the resistor is one-half of the original energy in C1. One-fourth of the energy is left in C1 and the other one-fourth is in C2.

For Case B, assume that R2 is zero. Fig 3 illustrates how the voltage V_{C1} and V_{C2} vary over the first cycle at the resonant frequency.

At the instant S2 is closed, a path for current flow is completed. The inductor and two capacitors form a series-resonant circuit. Therefore, when S2 is closed, a sine shaped current wave starts to flow. This current flow takes energy from C1 and stores it in L and C2. With equal capacitances, V_{C2} rises the same amount as V_{C1} decreases. The instantaneous energy in L is $(1/2)LI^2$. The current increases until $V_{C2} = V_{C1}$. At this time, t/4, the current wave peaks, 75% of the energy has been taken from C1, 50% has been stored in L and 25% is in C2.

Then the magnetic field in L starts to collapse and generates the voltage necessary to keep the current flowing for the next quarter cycle ending at t/2. At this point, the current has decreased to zero and all of the energy has been removed from C1 and stored in C2. The current flow then reverses and during the next half cycle all of the energy is transferred back to C1.

An interesting way to look at this behavior mathematically is as follows: At the instant S2 is closed, the dc voltage V on C1 instantly becomes the sum of two equal voltages. Half remains a dc voltage and the other half becomes the peak of an ac voltage.

$$V = \frac{V}{2} + \frac{V}{2}\cos(\theta) \text{ with } \theta = 0 \qquad \text{(Eq 1)}$$

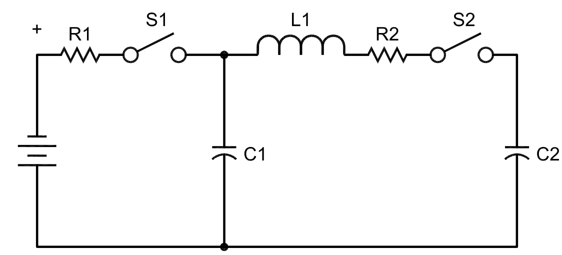


Fig 1—Circuit used for analysis.

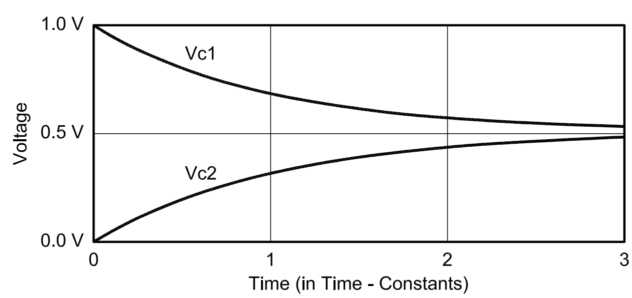


Fig 2—Voltages across capacitors for three time-constant periods.

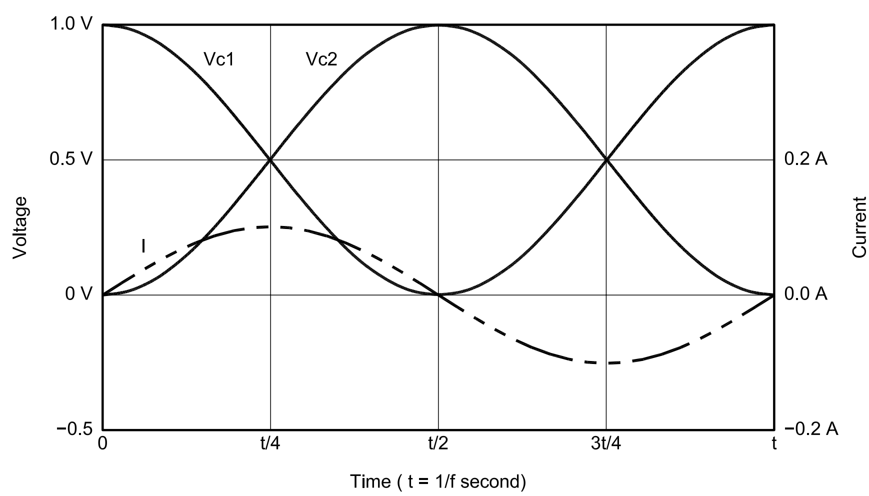


Fig 3—Voltage across capacitors during first cycle and the series-resonant current. Voltage across inductor is $V_{C!} - V_{C2}$.

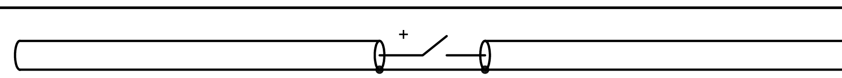


Fig 4—Diagram for computing dc energy transfer loss with distributed L and C.

The zero voltage on C2 instantly becomes the difference between two equal voltages.

$$\theta = \frac{V}{2}0 - \frac{V}{2}\cos(\theta) \text{ with } \theta = 0 \qquad \text{(Eq 2)}$$

The phase angle θ is a function of time and starts at zero the instant S2 is closed. The phase angle is 360° at the end of the first cycle of the resonant frequency.

The peak amplitude of the current sine wave can be computed from the energy in L at t/4 and the inductance of L.

In practice, there is always some loss resistance, which will cause the oscillation to decay exponentially. Converting the series resistance to the equivalent parallel resistance, R3, across the circuit, gives a value that can be used to compute the time-constant.

The value of C used to compute the time constant is the equivalent value of C1 and C2 in series. This value in Farads is used in the equation

$$T = \frac{R3C1}{2} \qquad \text{(Eq 3)}$$

The number of cycles in one time-constant period is equal to the *Q* of the resonant circuit.

$$Q = \frac{R3}{X_{C1} + X_{C2}} \qquad \text{(Eq 4)}$$

If *Q* were 100, it would take 100 cycles for the amplitude to decay to 37% of the original ac voltage.

$$\frac{1}{e} = \frac{1}{2.71828} = 36.78\% \qquad \text{(Eq 5)}$$

The peak amplitude of the ac (or RF) voltage decays as

$$V_{AC} = \left(\frac{V}{2}\right)e^{-\frac{t}{T}} \qquad \text{(Eq 5)}$$

The ac voltage is centered on the dc voltage of *V*/2.

It is observed that the positive and negative peaks of the ac voltage fall on the same voltage curves shown in Fig 2.

It takes about four time-constant periods, 4T, for the amplitude to die down to less than 1% of the original value. This happens very quickly. For a resonant frequency of 160 kHz and a *Q* of 100:

$$t1 = 4\frac{Q}{f} = \frac{4(100)}{160{,}000} = \frac{1}{400}s \qquad \text{(Eq 6)}$$

The circulating energy would be only $(0.01)^2$ or 0.0001 of the original value.

Thus, one-half of the original energy, *J*, in C1 is quickly dissipated, leaving the other half divided between dc charges in C1 and C2.

Case C: Instead of lumped capacitance and inductance, consider distributed capacitance and inductance in the form of transmission lines (see Fig 4). First we charge the capacitance of the left line with a voltage *V*. Then close S2. Wave fronts are launched in

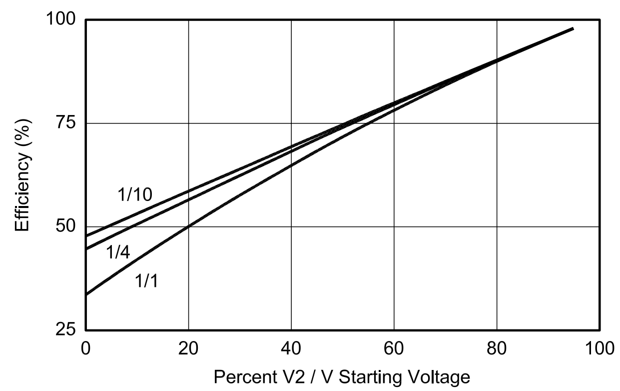


Fig 5—Efficiency versus *V2* starting voltage for C2 equal to C1, C1/4 and C1/10.

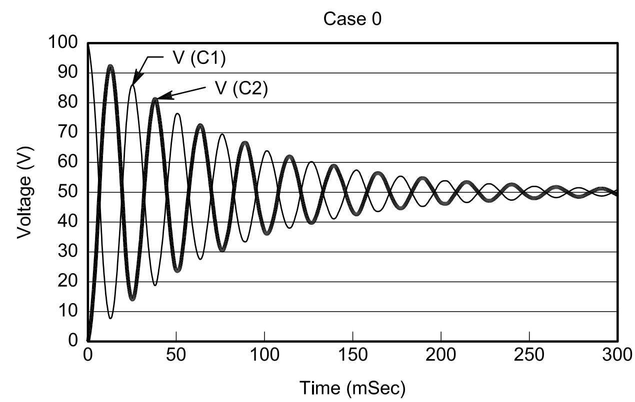


Fig 6—Typical damping action.

Table 1
Summary

	Case 0 no diode	*Case 1*	*Case 2*	*Case 3*	*Case 4*
C1 (µF)	8	8	8	8	8
C2 (µF)	8	8	8	2	8
L (H)	4	4	0.04	0.04	0.04
R(Ω)	100	100	2	4	2
Initial V(C1)	100	100	100	100	100
Initial V(C2)	0	0	0	0	50
Final V(C1)	50	7.792	2.376	60.98	51.47
Final V(C2)	50	91.23	97.52	155.2	98.32
Initial E(C1) (mJ)	40	40	40	40	40
Initial E(C2) (mJ)	0	0	0	0	10
Final E(C1) (mJ)	10	0.243	0.023	14874	10.597
Final E(C2) (mJ)	10	33.292	38.041	24.087	38.667
Lost E (rnJ)	20	6.465	1.937	1.039	0.736

opposite directions and with opposite polarities. These waves are reflected at the open ends of the lines. This sets up a half-wave resonance in the combined lines. A dc voltage of *V*/2 remains on both lines. Again, half of the original voltage in the left line remains as a dc charge in the combined distributed capacitance of both lines. The other half is dissipated as the half-wave resonance decays.

For laboratory observation of this decaying voltage using power-supply filter capacitors, a dc filter choke can be inserted in series with the switch to slow the resonant frequency to the low audio region. The loss resistance would be little more than the dc resistance of the filter choke.

The above analysis solves the mystery of where half of the energy goes. It is dissipated in a resonant circuit.

The efficiency of energy transfer can be improved some by using a value of C2 less than C1. This improvement is shown at the left edge of the diagram in Fig 5. More improvement can be realized when C2 is partly charged at the start. This improvement for three values of C2 is shown in Fig 5.

Now let us examine a means to move all of the energy in C1 to C2 and keep it there with "no" loss. First, add a lossless series inductor that resonates the capacitors in the audio region. Then add an ideal diode in series also. Now charge C1 and close S2. Fig 3 shows how *V1* and *V2* vary. The diode prevents reversal of the current, at t/2, which stops any further variation, leaving C1 with no charge and C2 charged to voltage V. Thus all of the energy has been transferred from C1 to C2.

In practice there is some energy dissipated in the energy transfer. There is loss resistance in series with the inductor and there is a small voltage drop across the diode, which dissipates some energy also. JB Jenkins, W5EU, ran a test using two 8 μF capacitors and a 4 H filter choke, which had 100 Ω of series resistance. The diode was a 1000 V, 2.5 A rectifier diode. C1 was charged to 30 V dc. After S2 was closed, C1 was left with 3 V representing approximately 1% of the original energy. There was 27 V on C2, which indicates that 81% of the original energy was transferred. The difference of 18% represents energy dissipated. He also verified that without the inductor and diode the voltage on both capacitors was 15 V, which was one half of the original 30 V charge. This test verified my theory.

For more accurate, computer generated data, use a transient simulation program called *SPICE*. This provides an easy way to try different values of components and starting voltage charges on C2. Roy Culbertson, AD5EQ, volunteered to run a few circuit variations for further verification of my theory and to establish a few properties of the circuit.

Case 0: This represents the case with equal 8 μF capacitors, a 4 H filter choke with 100 Ω resistance but no diode. An initial charge of 100 V

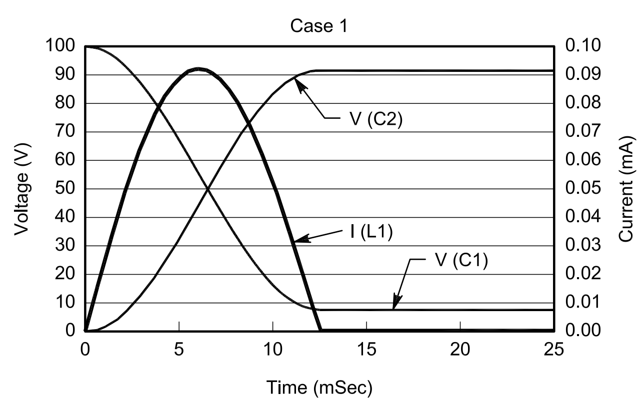


Fig 7—Case 1: equal 8 μF capacitors, a 4 H filter choke with 100 Ω resistance but no diode.

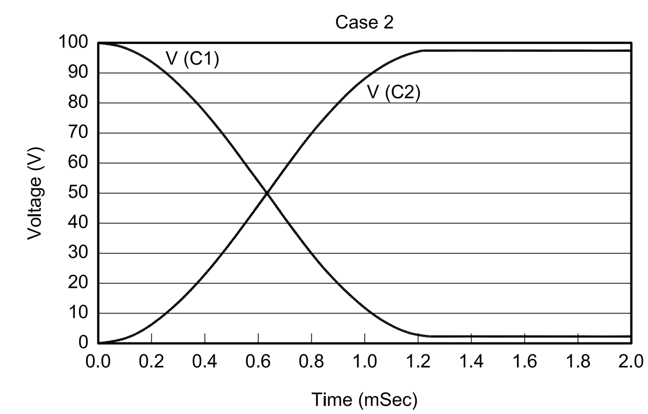


Fig 8—Case 2: the inductor was reduced to 0.04 H providing a resonant frequency of 400 Hz. A resistance of 2.0 Ω was chosen to provide a Q of 50.

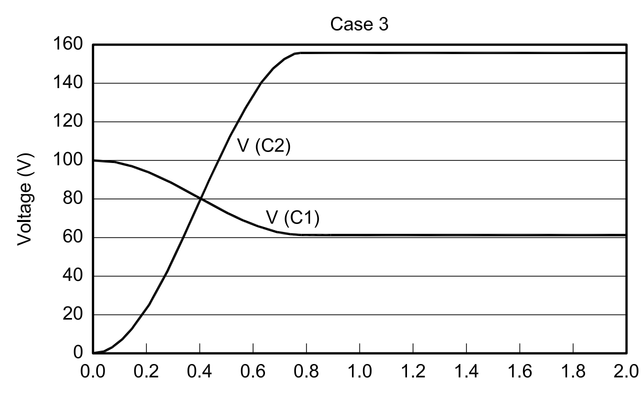


Fig 9— Case 3*:* C2 was reduced to 2 μF.

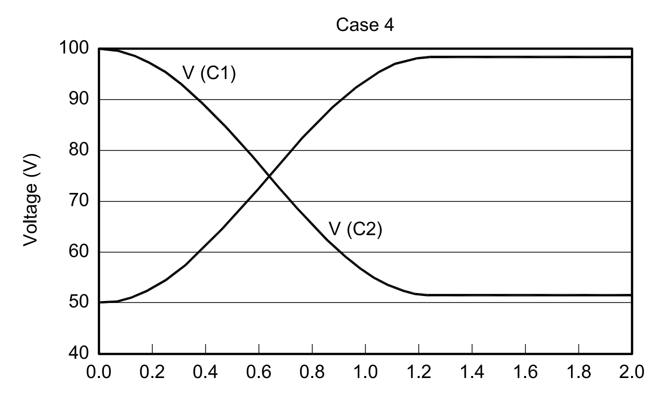


Fig 10— Case 4: Equal capacitors were used but C2 was initially charged to 50 V, which is half of the 100 V charge on C1.

was used. After closing S2, the voltage on each capacitor was 50 V.

Case 1: The parameters were chosen to approximate the values used by W5EU in his lab. The two capacitors were 8 µF each, the inductor was 4 H with 100 Ω resistance. The resonant frequency was 40 Hz, the inductive reactance was $j1000$ and the inductor Q was 10. The result is shown in Fig 7, and a tabulation of data obtained from the *SPICE* program is shown in Table 1.

Case 2: The inductor was reduced to 0.04 H providing a resonant frequency of 400 Hz. A resistance of 2.0 Ω was chosen to provide a Q of 50. Fig 8 and Table 1 show the results. Over 95% of the energy was transferred with only 5% dissipated. Less than 0.1% was left in C1.

Case 3: C2 was reduced to 2 µF. This raised the resonant frequency to 515 Hz and an inductive reactance of $j193$ Ω. A resistance of 4 Ω was used for a Q near 50. Fig 9 shows the result. Notice that 60% of the initial energy was transferred with a 55% voltage stepup. Only 4.1% of the energy removed from C1 was dissipated.

Case 4: Equal capacitors were used but C2 was initially charged to 50 V, which is half of the 100 V charge on C1. The initial voltage charges were close to being reversed as only 2.5% of the energy removed from C1 was dissipated. See Fig 10 and Table 1.

I hope that this information has started the "wheels turning" in some readers' minds, which will lead to improved designs. These tests and analysis were just enough to prove the theory. Quite likely still better transfer efficiency can be realized by using the optimum value of inductance and a higher Q inductor. The current increases as the inductance is decreased. This causes the diode loss to increase. The *SPICE* program would be a valuable aid for optimizing a proposed circuit.—*Warren Bruene, W5OLY, 7805 Chattington Dr, Dallas, TX 75248-5307*

Evaluation of Antenna Tuners and Baluns (Sep/Oct 2003)

There is an error in Frank Witt's, AI1H, article in the Sep/Oct 2003 *QEX*. Fig 4 (on page 6) should have exactly 67 points on the line of reflection-coefficient magnitudes that can be displayed by the MFJ-259B.

In addition, the author has discovered an error on page 10, left column, in the first full paragraph. The error bounds are the same for halving R_L as for doubling R_L. So, if the loss were calculated without using geometric averaging (by either halving or doubling R_L), the calculated value could be too high by as much as 10.8% or too low by as much as 16%. Halving or doubling R_L can yield either positive or negative errors, and the error bounds are identical for the two cases.

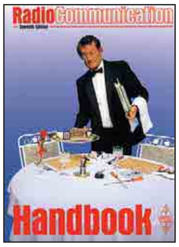
The corrected article has been placed on the ARRLWeb at **www.arrl.org/tis/info/pdf/030910qex003. pdf**.—*Bob Schetgen, KU7G,* QEX *Managing Editor;* **ku7g@arrl.org** ◻◻

Next Issue in QEX/Communications Quarterly

Brain Cake, KF2YN, begins a two-part series on twin-C and box-kite antennas. We shall take them separately, although they are related in a way. Also, Bob Kopski, K3NHI, presents a calibration modification to his "Advanced VHF Wattmeter." ◻◻

RSGB PRODUCTS

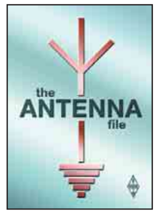
Imported by ARRL— from the Radio Society of Great Britain



Radio Communication Handbook

One of the most comprehensive guides to the theory and practice of Amateur Radio communication. Find the latest technical innovations and techniques, from LF (including a new chapter for LowFERS!) to the GHz bands. For professionals and students alike. 820 pages.

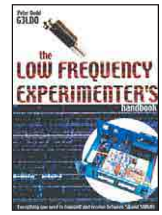
ARRL Order No. 5234—**$53**



The Antenna File

The best work from the last ten years of RSGB's *RadCom* magazine. 50 HF antennas, 14 VHF/UHF/SHF, 3 on receiving, 6 articles on masts and supports, 9 on tuning and measuring, 4 on antenna construction, 5 on design and theory. Beams, wire antennas, verticals, loops, mobile whips and more. 288 pages.

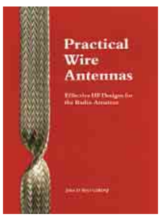
ARRL Order No. 8558—**$34.95**



The Low Frequency Experimenter's Handbook

Invaluable reference and techniques for transmitting and receiving between 50 and 500 kHz. 112 pages.

ARRL Order No. RLFS—**$32**



Practical Wire Antennas

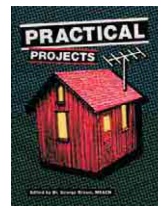
The practical aspects of HF wire antennas: how the various types work, and how to buy or build one that's right for you. Marconis, Windoms, loops, dipoles and even underground antennas! The final chapter covers matching systems. 100 pages.

Order No. R878—**$17**

Antenna Toolkit 2

The complete solution for understanding and designing antennas. Book includes a powerful suite of antenna design software (CD-ROM requires *Windows*). Select antenna type and frequency for quick calculations. 256 pages.

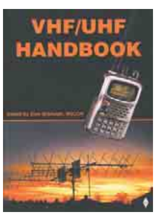
ARRL Order No. 8547—**$43.95**



Practical Projects

Packed with 50 simple "weekend projects." A wide variety of radio and electronic ideas are covered, including an 80-m transceiver, antennas, ATUs and simple keyers.

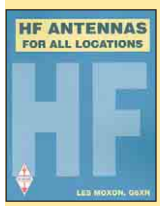
ARRL Order No. 8971—**$24.95**



VHF/UHF Handbook

The theory and practice of VHF/UHF operating and transmission lines. Background on antennas, EMC, propagation, receivers and transmitters, and construction details for many projects. Plus, specialized modes such as data and TV. 317 pages.

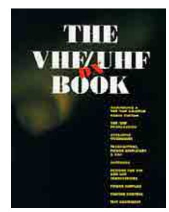
ARRL Order No. 6559—**$35**



HF Antennas for All Locations

Design and construction details for hundreds of antennas, including some unusual designs. Don't let a lack of real estate keep you off the air! 322 pages.

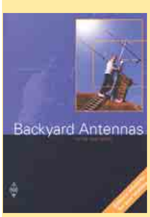
ARRL Order No. 4300—**$34.95**



The VHF/UHF DX Book

Assemble a VHF/UHF station, and learn about VHF/UHF propagation, operating techniques, transmitters, power amplifiers and EMC. Includes designs for VHF and UHF transverters, power supplies, test equipment and much more. 448 pages.

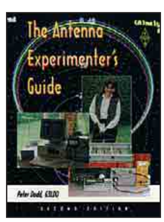
Order No. 5668—**$35**



Backyard Antennas

With a variety of simple techniques, you can build high performance antennas. Create compact multi-band antennas, end-fed and center-fed antennas, rotary beams, loops, tuning units, VHF/UHF antennas, and more! 208 pages.

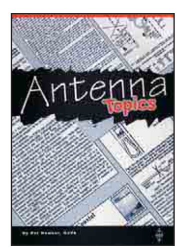
ARRL Order No. RBYA—**$32**



The Antenna Experimenter's Guide

Build and use simple RF equipment to measure antenna impedance, resonance and performance. General antenna construction methods, how to test theories, and using a computer to model antennas. 158 pages.

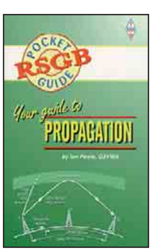
ARRL Order No. 6087—**$30**



Antenna Topics

A goldmine of information and ideas! This book follows the writings of Pat Hawker, G3VA and his "Technical Topics" column, published in *Radcom*. Forty years of antenna design.

ARRL Order No. 8963—**$34.95**



Your Guide to Propagation

This handy, easy-to-read guide takes the mystery out of radio wave propagation. It will benefit anyone who wants to understand how to get better results from their station.

ARRL Order No. 7296—**$17**

Guide to EMC #7350 **$34**

IOTA Directory—11th Edition #8745 **$16**

NEW! Microwave Projects #9022 **$26**

NEW! QRP Basics #9031 **$26**

Radio & Electronics Cookbook #RREC **$28**

NEW! RSGB Prefix Guide —6th Edition #9046 **$16**

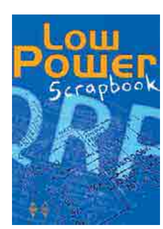
Technical Compendium #RTCP **$30**

Technical Topics Scrapbook

1985-1989 edition #RT85 **$18**

1990-1994 edition #7423 **$25**

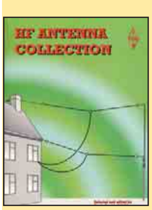
1995-1999 edition #RT95 **$25**



Low Power Scrapbook

Build it yourself! Low power transmitters, simple receivers, accessories, circuit and construction hints and antennas. Projects from the G-QRP Club's magazine Sprat. 320 pages.

ARRL Order No. LPSB—**$19.95**



HF Antenna Collection

Articles from RSGB's *RadCom* magazine. Single- and multi-element horizontal and vertical antennas, very small transmitting and receiving antennas, feeders, tuners and more. 240 pages.

ARRL Order No. 3770—**$34.95**

Order Toll-Free
1-888-277-5289
www.arrl.org/shop

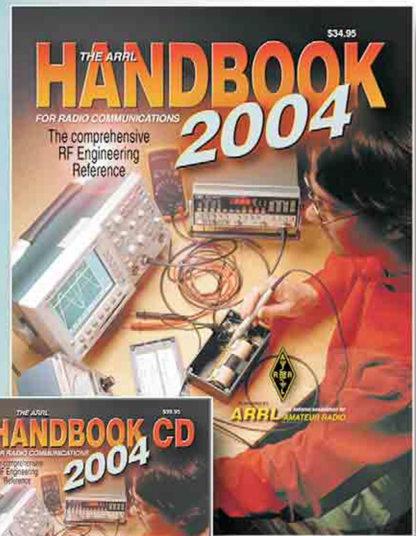
Shipping: US orders add $5 for one item, plus $1 for each additional item ($10 max.). International orders add $2.00 to US rate ($12.00 max.). US orders shipped via UPS

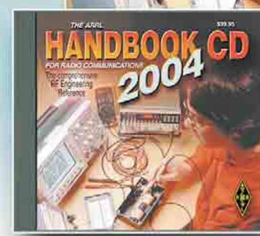
ARRL *The national association for* ***AMATEUR RADIO***

225 Main Street • Newington, CT 06111-1494 USA

tel: 860-594-0355 fax: 860-594-0303
e-mail: pubsales@arrl.org
www.arrl.org/

"I can build it."





**Some supplementary software utilities included—for *Windows* and DOS only.

New Eighty-First Edition

The ARRL Handbook for Radio Communications—2004

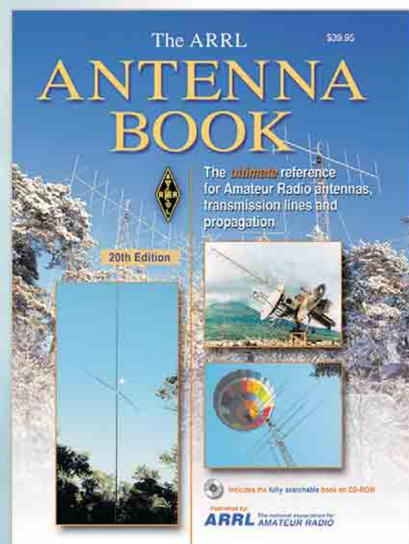
The Standard in applied electronics and communications! Filled with projects, antennas, and indispensable references. Always revised. Now including a commercial-quality, high-voltage power (plate) supply, and a revised version of a high-performance, easy-to-build passive CW filter.

Softcover
ARRL Order No. 1964 **$34.95***

Hardcover
ARRL Order No. 1972 **$49.95***

The ARRL Handbook CD for Radio Communications
Version 8.0—for *Windows* and Macintosh**
View, Search and Print from the entire 2004 edition book!

CD-ROM
ARRL Order No. 1980 **$39.95***

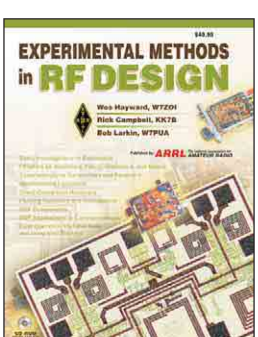


New Twentieth Edition

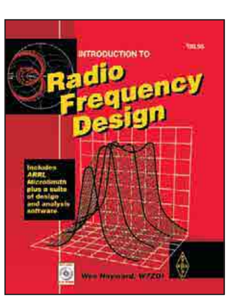
Hams rely on **The ARRL Antenna Book** for current antenna theory and a wealth of practical, how-to construction projects. Extensively revised, and featuring antenna designs enhanced by the latest advances in computer modeling. **Includes the fully-searchable book on CD-ROM** and additional software utilities.

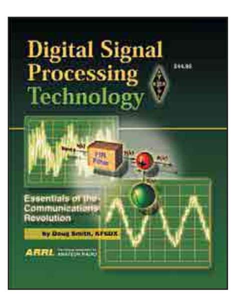
Book with CD-ROM
ARRL Order No. 9043 **$39.95***

Try these other ARRL Publications



Experimental Methods in RF Design
ARRL Order No. 8799 ...... **$49.95***





Digital Signal Processing Technology—Essentials of the Communications Revolution
ARRL Order No. 8195 .... **$44.95***

Introduction to Radio Frequency Design includes software.
ARRL Order No. 4920 .... **$39.95***

***Shipping:** in the US, add the following amounts to your order (UPS). An additional $2.00 will be added to the US rate for shipment outside the US (surface). Air delivery and other shipping services are available.

Amount of Order	Add
$30.01 to $40.00	$7.00
$40.01 to $50.00	$8.00
$50.01 to $75.00	$9.00
Over $75.00	$10.00
CD-ROM only	$5.00

Sales tax is required for orders shipped to CA, CT, VA, and Canada. Prices subject to change without notice.

ARRL *The national association for AMATEUR RADIO* **225 Main Street, Newington, CT 06111-1494 tel: 860-594-0355 fax: 860-594-0303**

In the US call our toll-free number **1-888-277-5289** 8 AM-8 PM Eastern time Mon.-Fri. **www.arrl.org/shop**