Conference Proceedings

RFexpo EAST

October 19-21, 1993
Tampa Convention Center
Tampa, Florida

Sponsored by RFdesign Magazine

Conference Proceedings

RF expo EAST

October 19-21, 1993
Tampa Convention Center
Tampa, Florida

Argus Business
6300 S. Syracuse Way
Suite 650
Englewood, Colorado 80111
(303) 220-0600

Sponsored by RF design Magazine

TABLE OF CONTENTS

An ISM Band Design for WLAN and PCS

Robert J. Zavrel
Worldwide Applications Manager, WLAN Products
GEC Plessey Semiconductors
1500 Green Hills Road
Scotts Valley CA 95066

1. Introduction

To realize the promise of personal wireless communications on a massive scale, a complex set of business, political, and technical issues must be resolved. New frequency allocations seem to be a certainty within the next few years. However, by using the existing ISM bands reliable and economic systems can now be configured addressing very large volume markets. This paper presents the design methodology for such a system.

The three Industrial, Scientific, and Medical band allocations (ISM bands) in North America are quite attractive for a wide range of would-be wireless services. These bands not only pose the opportunity but also design challenges to realize effective systems. FCC regulations permit up to one watt of RF output power without requiring a license. In order to facilitate *some degree* of interference rejection in a license-free environment, spread spectrum techniques are mandated. The two types of spread spectrum techniques that are permitted are frequency hopping and direct sequence. Much has been written over the past few years about these two techniques and the reader can seek the references for background information. (Ref. 1 & 2).

2. Architecture Considerations

In order to optimize an architecture one of three bands and one of two spread spectrum techniques must be chosen. A complex set of system trade-offs must be considered in this decision-making process. A review of the status of the three bands, together with design considerations can lead to a rational choice.

The 902 MHz band was recently labeled "the kitchen sink band" (Ref. 3). This is indeed an accurate description. Multiple users include radiolocation services, amateur radio, fixed and mobile services and AVM. In addition, the FCC has proposed wind profile radars and AVM for the entire band. Furthermore, the 902 MHz band is only available in North America limiting its desirability for most high volume manufacturers. This band, however, has the advantage that silicon devices can be used efficiently in both the receiver front ends and transmitter PAs. Use of this band for PCS and/or WLANs is tenuous at best.

The 2.4 GHz band is partially allocated to amateur radio. Microwave ovens also operate in this band. However, compared with the 902 MHz band, these users are comparatively benign. Another major advantage of the 2.4 GHz band is that allocations exist within the 2.4 to 2.5 GHz band in North America, Europe, and Japan. For these reasons the 2.4 GHz band was chosen for our design solution.

The 5.8 GHz band is also attractive, but the higher operating frequency offers greater design and manufacturing problems and higher power consumption. This band might be quite interesting for future generation products, if exclusive PCS/WLAN bands are not allocated in the near future.

Having chosen the 2.4 GHz ISM band the question of which spread spectrum technique remains. There are two basic spread spectrum techniques: direct sequence and frequency hopping. In contrast to direct sequence systems, frequency hopping systems offer better immunity to interference. In short, a direct sequence system relies totally upon processing gain (and the resulting jamming margin specifications) to reject in-band interference.

$$G(p) = 10 \log Bs/Rd$$

where G(p) is the processing gain, Bs is the spreading bandwidth, and Rd is the data rate and

$$Mj = G(p) - (L + (S/N))$$

where Mj is the jamming margin, L is system losses, and S/N is the signal-to-noise ratio necessary to maintain a given bit error rate (BER) or error probability (Pe).

These equations are fundamental to direct sequence systems. Indeed, even low data rate direct sequence systems mandate the use of sophisticated power level control of all transmitters operating in band. The unpredictability of the ISM band environment typically renders direct sequence systems unreliable for use in these bands. There may be some special circumstances where direct sequence can be the technique of choice, but the interference trade-off is omnipresent. Perhaps the most important mitigating circumstance is for data rates exceeding about 2 Mb/sec. It is difficult to achieve over 2 Mb/sec. with frequency hoppers while maintaining the initial advantages of the hopping system. However, as data rates increase in the direct sequence system, the vulnerability also increases. Furthermore, for each doubling of data rate, the system noise figure increases by 3 dB resulting in a net reduction in range. These might be an acceptable trade-offs, particularly if communications is intended for very short ranges, perhaps under 3 meters. However, for most practical applications, the data rate in ISM applications is limited to about 2 Mb/sec.

Frequency hopping systems rely upon the more traditional techniques of radio receiver design: dynamic range, sensitivity and selectivity. These basic specifications, in turn, reflect such specifications as noise figure, local oscillator spectral purity, filter skirts, and intermodulation. In contrast, in-band selectivity in direct sequence systems relies completely upon processing gain.

For example, in a typical ISM direct sequence system, 5 dB of jamming margin is typical (for 800 kb/sec data rates, 10E-5 BER, operating in the US 2.4 GHz ISM band). In contrast, frequency hopping systems easily provide 60 dB of ultimate rejection by using cascaded SAW filters. Under some circumstances the LO phase noise can become more constraining than the IF filter skirts. In frequency hopping systems, a narrow-band interferer might "wipe out" several hopping channels. With proper coding, however, individual "packets" can be recovered despite the loss of a considerable number of hopping channels (Reference 4). In the direct sequence case, if the interferer exceeds the jamming margin of the system, the bit error rate (BER)

begins to increase. Further reduction in interference will upset the link and communications will be lost completely rather than a percentage of packets. Some direct sequence systems combine frequency channelization to provide filter selectivity in addition to processing gain. However, it is our conclusion that once frequency division schemes must be employed, the designer might as well limit bandwidth to the baseband and realize the full advantages narrow-band systems offer. In other words, use frequency hopping.

Direct sequence systems can be very attractive *only* in a pristine wide band where all in-band transmitters are rigidly controlled. Direct sequence systems **must** control all transmitters' power output in real time, or one or more transmitters are likely to interfere with each other. Thus, direct sequence systems must assume that all users of the spectrum within range will conform to that system's control mechanisms. If any signal source appears in-band that exceeds the jamming margin of a receiver, the bit error rate or the link itself will be compromised. Of course, this premise is not possible to attain within the ISM bands. Indeed, even with exclusive use of a band complex control mechanisms are required to control the real-time power outputs of the transmitters. Furthermore, these control mechanisms compromise the power consumption and price of the system. The ISM bands are anything but pristine. Indeed the ISM bands are hostile communications environments requiring excellent interference-avoidance techniques. The ISM band environment is not unlike circumstances presented to electronic warefare (EW) designers. It is interesting to note that tactical military communications systems typically use frequency hopping techniques to mitigate enemy jammers. Figure 1 shows in-band equivalent selectivity for the 83 MHz-wide 2.4 GHz band for various data rates of direct sequence systems contrasted against a frequency hopping system. The jamming margin numbers are based on 10 dB necessary C/N ratio for 10 E-5 BER and 5 dB system loss.

3. System Requirements

Having settled upon 2.4 GHz frequency hopping as the architecture of choice, a detailed design analysis can now proceed. In order to "open" the very high volume data and voice communications markets certain criteria must be met in the design solution:

Power consumption: about 350 mW in receive mode
Cost: under $100.00 for high volume (>500K/year)
Size: PCMCIA form factor (about 2x3 inch circuit board)
Performance:

A. Data Rate: up to 1 Mbit/second
B. Dynamic range: >60 dB (defined by filter skirts, noise figure, and compression point)
C. Dual antenna diversity
D. Receiver noise figure: <10 dB
E. double conversion superhet design for outstanding adjacent channel rejection
F. Tx-to-RX and RX-to-TX time <4 uS
G. Hopping settling time: <100uS
H. RF Power Output: selectable 10 or 100 mW.

The extremely fast transmit-receive transition time and hopping speed permits the use of state-of-the-art control algorithms and excellent acquisition times. Double conversion minimizes image response, permits the use of two SAW filters for outstanding selectivity, and permits limiting and detection at the optimum frequency.

Most of the end-use system requirements such as range, BER, data rate, and reliability reduce to the set of RF specifications listed above. These specifications, in turn, must conform to the power, cost and form factor requirements. These combined specifications reflect a fairly high performance microwave transceiver built on a credit card from full custom devices, selling for under $100.00. This challenge has been met with the GEC Plessey DE6003. A microwave transceiver on a credit card sized circuit board with the above set of specifications speaks for itself as a state-of-the art design.

4. RF Transceiver Description

The DE6003 represents the most advanced license-free radio data transceiver available for low power, low cost, miniature packaged and high performance applications. These transceiver modules are designed for portable battery-operated data or voice communications equipment. For example, lap-top and hand-held computers, cordless telephones, point-of-sale, inventory, security, and medical equipment can utilize these modules to great advantage. Figure 2 shows an interface block diagram of the DE6003 with power supply, antenna, and digital control bus.

Figure 3 shows an RF functional diagram with divisions of the integrated circuits delineated. Three PLL synthesizers run continuously to provide the very high hopping and TX/RX speeds. A patented divide-by-two switch is used in the transmit loop to prevent interference into the receiver IF. Dual conversion facilitates excellent selectivity and image rejection.

The DE6003 covers 2.400 to 2.500 GHz and utilizes a binary Gaussian frequency shift keying modulation scheme (GFSK). GFSK has several distinct advantages over alternative baseband schemes. The IF systems do not require linearity and the complications involving AGC design. Limiters are easy to design and manufacture and signal strength indication (RSSI) is also relatively easy to realize. In the hostile ISM environment with interference and multipath distortion, GFSK also provides a effective and simple solution. SImplicity, manufacturability, low cost, and low power consumption are all inherent characteristics of GFSK systems. The main disadvantage of GFSK is a comparatively low bit/Hz/sec specification (typically 0.7 bits/Hz/sec.). Also, M-ary FSK, particularly 4FSK can be used to increase data rates well above 1b/Hz/sec. and is being investigated for second generation products. However, most of the current high volume market applications require well under 1Mb/sec. rates, reflecting the DE6003 specifications.

Early in the design stage of the transceiver it was realized that the effort would necessitate full custom integrated circuits. During innumerable design reviews power consumption (milliwatt by milliwatt) and cost were traded-off the various performance requirements (data rate, dynamic range, speed and others). GEC Plessey's design effort was greatly enhanced in that the IC, filter, and system designers all report through the same program manager. Consequently, optimum designs were possible from the various processes used while maintaing tight scheduling. Aggressive pricing was realized because all the critical components for the transceiver are fabricated within the same company. The radio takes full advantage of the three most important semiconductor manufacturing processes: GaAs, high speed bipolar, and CMOS.

World-class design people and facilities exist within GEC Plessey's multiple facilities. (GEC Lincoln started

in the microwave business during World War II manufacturing point contact diodes for early radar sets.) Since then, the company has built a very impressive array of RF integrated circuits encompassing markets from consumer to space qualification. Traditional strengths have included synthesizers and low power RF integrated circuits, both critical to successful specification conformance.

The WLAN development program was broken into three stages:

1. Proof of concept (completed March, 1992)
2. System development "Alpha" units (completed September, 1992, DE6002)
3. Production prototype samples (available February 1994, DE6003)

"Beta" units of the DE6003 are currently being evaluated and design-ins are taking place in many companies.

The next generation of specifications will require MCM (multi-chip module) technology. The notion of a monolithic WLAN radio with the above specifications is currently not possible. Optimum performance must be realized by using optimum processes and designs for each function. There are no silicon processes currently available that meet all these requirements. Furthermore, Bi-MOS processes have limited advantages in meeting overall system objectives.

5. Digital Control Interface

Table 1 shows the details of the digital control interface. All interface functions to the DE6003 use standard CMOS levels. All the interface connections are digital except where otherwise noted. The digital control interface allows direct control of the radio's functions. The digital hardware that interfaces to the radio can be tailored to any specific application and any set of regulations. Consequently the DE6003 affords the system designer great flexibility for data, voice, and even video transmission systems. All the remaining circuitry to complete a data communications transceiver can be implemented in a CMOS ASIC.

TABLE 1

TXD: Serial data input to the transmitter
RXD: Serial data output from the receiver
SYNLOK: Indicates when all PLL synthesizers are locked
RADON: On/off switch for the radio
PLC: Power level control selects either 10 or 100 mW transmitter RF output
DSEL: Diversity select one of the two antennas
PAON: On/off switch for the PA transmitter stage
TXB/RX: Selects either transmit or receive modes
CLK: 10 MHz clock output (200 mV PTP)
RSSI: Received signal strength indicator (analog voltage proportional to dB received signal level).
LOADBAR: Loads frequency channel select
CHANNEL SELECT: Parallel seven bit word defines frequency used by the radio from 2.4 to 2.5 GHz.

6 . Antenna and EMI Considerations

Placing the antenna on the circuit board is not acceptable because the radio would require shielding to prevent unwanted radiation from the radio circuitry. Also, the optimum placement of the antenna is typically not the optimum place for the radio proper. Maximizing receiver and transmitter performance only to compromise the effort by poor antenna placement is an iladvised proposition. Most antenna problems can be minimized if external antennas are permitted. At 2.4

GHz external antennas can assume small profiles, with helictical configurations fitting into one or two inch plastic hemispheres.

Incorporation of these microwave radio transceivers into portable computer-based hardware will require some new approaches to hardware design. Electromagnetic considerations, up to now, have been limited to electromagnetic interference (EMI) issues in computer hardware design. Extensive shielding is required to keep unintentional RF noise produced by high speed computer clocks from escaping the computer and interfering with radios, televisions, and other equipment. With radios and antennas "inside" the electromagnetic considerations become far more complex. Indeed, shielding will remain critical, but at the operating frequencies (2.4 GHz) the "unit" must provide the opposite function: transmit and receive antenna. Indeed, electromagnetic issues necessitate new ways of thinking about computer product engineering. Handheld and laptop microprocessor based equipment must now be considered complex electromagnetic devices.

7. Protocols, control, and standards

It is one problem to build an optimum radio for the above given constraints. It is another problem to control the radio, control the communications link, provide for clock recovery, synchronization, and interface to the data bus and networks. This is the work of the protocol circuit. GEC Plessey is developing several protocols and is providing interfaces to many of the important data communications standards: ISA, ETHERNET, PCMCIA, and others. Much of this work is covered by NDAs (non-disclosure agreements) in place with major customers. Consequently details cannot as yet be published. Suffice it to say that the protocol and interface considerations are not trivial and that a complete solution (antenna to data communications standard socket) will be available in early 1994. However, the key to opening the ISM and other radio bands is to build adequate radios.

The entire question of protocols begs the introduction and adoption of standards for WLANs. These standards will have to embrace not only ISM band radios but also IR and other radio band systems as well. The *IEEE 802.11* group and the new *IEEE Communications Society Technical Committee on PCS* are working on such standards. In addition, ETSI is working on European standards and *WINForum* (an industry group) is dealing with securing new bands and writing an etiquette for these bands' use. Clearly, convergence on standards will occur within the next year, either through these and other standards groups or by de-facto standards.

8. Conclusions

Efficient, reliable, and cost effective wireless digital communications systems are now possible to address the high volume portable markets. The opening of this market has been accomplished by providing good microwave transciever specifications derived from exhaustive market analysis and design processes. Reliable use of the difficult ISM environment has been accomplished with designs that can be translated to new WLAN/PCS frequencies quickly when these bands become available. Effective use of these radios requires interface CMOS circuits curently under development. Standards must also be set and accepted. Electromagnetic issues for antennas and shielding must also be rectified. A comprehensive program to address WLAN and PCS radio applications is in place at GEC Plessey Semiconductors. Standard

products that solve all problems between the data communications interface and the antenna will soon be available on the open market.

9. References

1. *Spread Spectrum Systems,* Robert C. Dixon, John Wiley & Sons, 1984

2. *Spread Spectrum Communications, Volumes 1, 2, & 3.* Marvin K. Simon, Jim K. Omura, Robert A. Scholtz, Barry K. Levitt, Computer Science Press, 1985

3. *The Kitchen-Sink Band: 902 MHz, "It Seems to Us...", QST,* David Sumner, American Radio Relay League, June, 1993

4. *Wireless Access Methods and Physical Layer Specifications,* Ed Geiger, IEEE 802.11 Document number IEEE P802.11-93/104, July, 1993

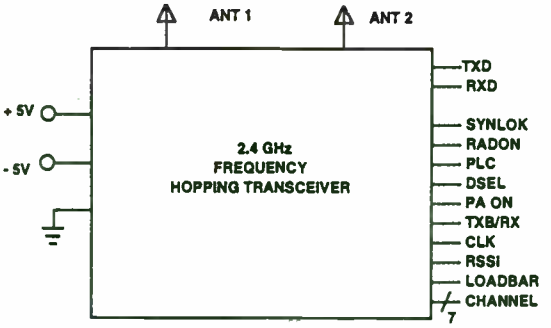


FIGURE 2
Functional Block Diagram of DE6003

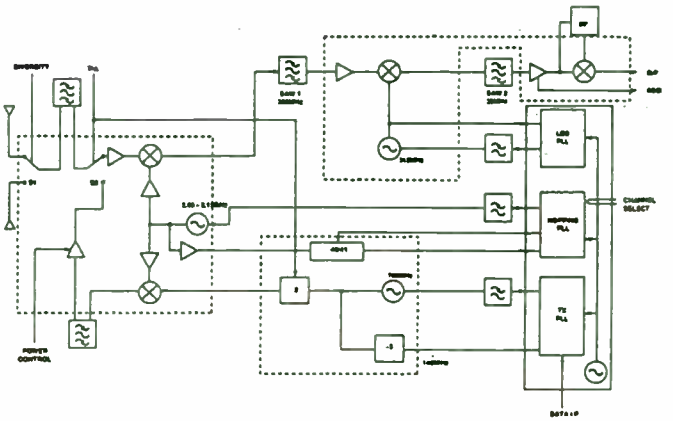


FIGURE 3
Detailed Block Diagram of DE6003

IN-BAND EQUIVALENT SELECTIVITY
FOR FREQUENCY HOPPING AND VARIOUS
DIRECT SEQUENCE SYSTEMS FOR 10^{-5}BER

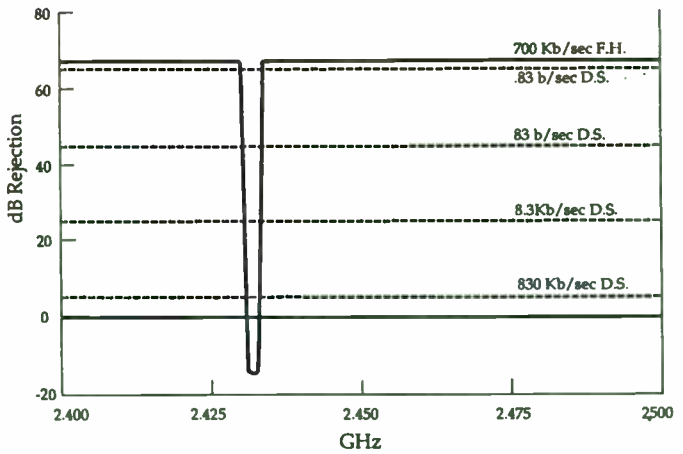


FIGURE 1.

A DSP Microprocessor Based Receiver for a Cosine Transition-shaped BPSK Signal

Roy E. Greeff and Bruce H. Williams
Unisys Government Systems Group
640 N. 2200 W. MS-F1E03
Salt Lake City, UT 84116

In a previous paper (RF EXPO East 92) we discussed a cosine transition-shaping BPSK modulator implemented by direct digital synthesizers. This waveform has advantages over conventionally filtered BPSK in that filter distortion effects can be tightly controlled and reproduced consistently This paper continues with the development of a system utilizing the shaped BPSK signal. This paper discusses a DSP microprocessor receiver for the cosine transition-shaped BPSK signal.

Modulation Waveform

By shaping NRZ data transition shapes one can effectively filter the spectrum. Figure 1 shows an "eye-pattern" for the shaped NRZ signal.

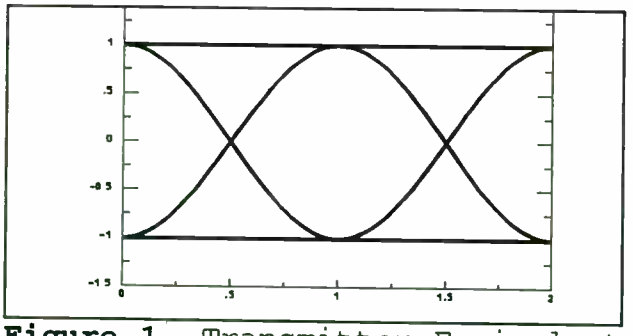


Figure 1 Transmitter Equivalent Eye Pattern

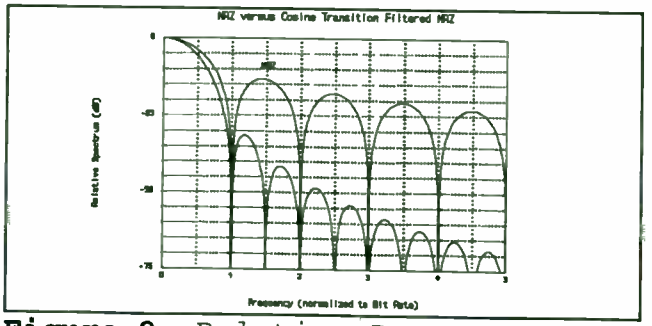


Figure 2 Relative Power Spectral Densities (dB/Hz)

Figure 2 demonstrates the effect of cosine transition shaping on the spectrum. Energy in the sidelobes is significantly reduced.

The receiver necessary to optimally demodulate this signal differs from conventional BPSK demodulators in that the matched filters are combined with intersymbol interference (ISI) canceling filters.

Main Receiver Functions

Figure 3 shows a block diagram of the main receiver functions necessary for PSK demodulation.

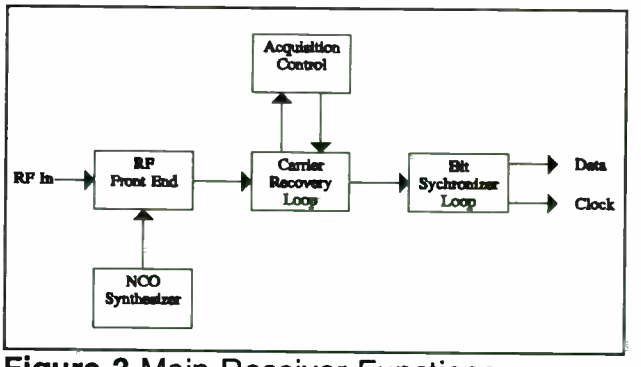


Figure 3 Main Receiver Functions

The RF front end consists of filters, the AGC, downconverter, and the A/D converters. The rest of the functions are implemented using DSP products. These functions in the past have been implemented using analog circuits. With the advent of higher speed DSP products (such as the Texas Instruments TMS320C40) these functions can now be implemented digitally for low data rates. The receiver built at UNISYS implements the Data Recovery Loop, acquisition control, and the bit

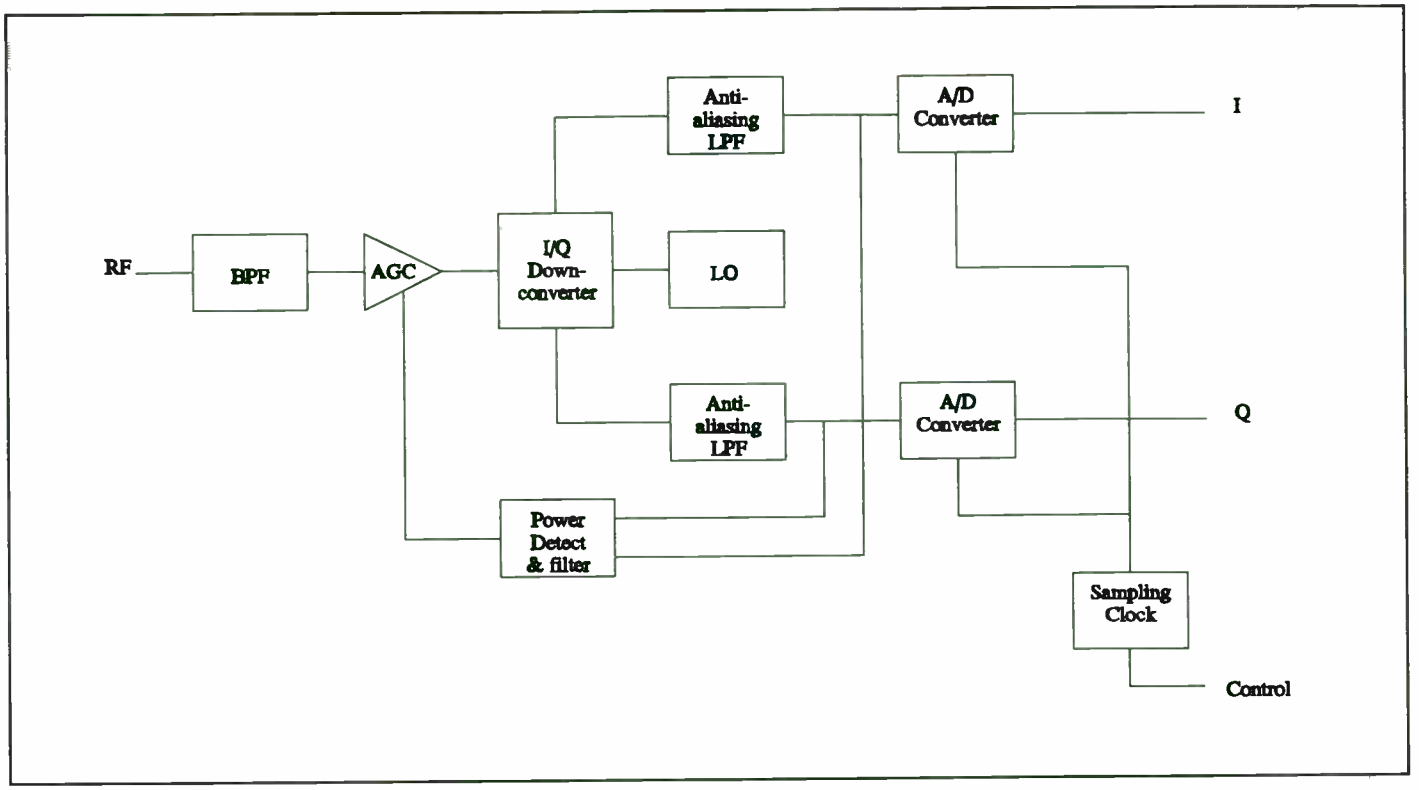


Figure 4 RF Front End

synchronizer with software loaded on a TI C40 floating point processor. Considerable flexibility is obtained by programming these functions into the C40. This flexibility is most apparent in the different algorithms that can be programmed (with no hardware changes) into the C40 for different modulation types (PSK, FSK, etc.).

RF Front End

Figure 4 shows a block diagram of the RF font end of the receiver.

The first filter is used to limit the noise bandwidth of the receiver. It's bandwidth is chosen to be only wide enough for the tuning flexibility required for multiple channels. The AGC amplifiers follow the filter. The detection of power for the AGC is at the output of the anti-aliasing filters. The power detection could occur after the baseband matched filters but if this is done to much power may be input into the I/Q downconverter. Additional AGCing is done at baseband in the demodulator algorithm. After the AGC amplifiers, the signal is downconverted as close to baseband as the frequency uncertainty of the received signal. An NCO is used as the synthesizer for the downconversion. The resultant I and Q signal is then filtered with the anti-aliasing filters. These signals are converted to digital signals using A/D converters operating at four times the maximum data rate. The sampling rate oscillator is controlled by the demodulator algorithm. The signals are fed to the C40 processors where the samples are used to perform the appropriate demodulation. The RF circuitry can be used with many different data rates and modulation types with no changes to the hardware.

Software Carrier Recovery Loop

Using DSP techniques to perform the demodulation vs analog techniques has various advantages. A demodulator design using analog circuits may degrade over temperature, vibration or other environmental factors. The software loaded into a digital processor is immune to environmental factors. A software demodulator can be changed easily to accommodate new data rates or even new modulation types. The one advantage that analog circuits have over the same digital

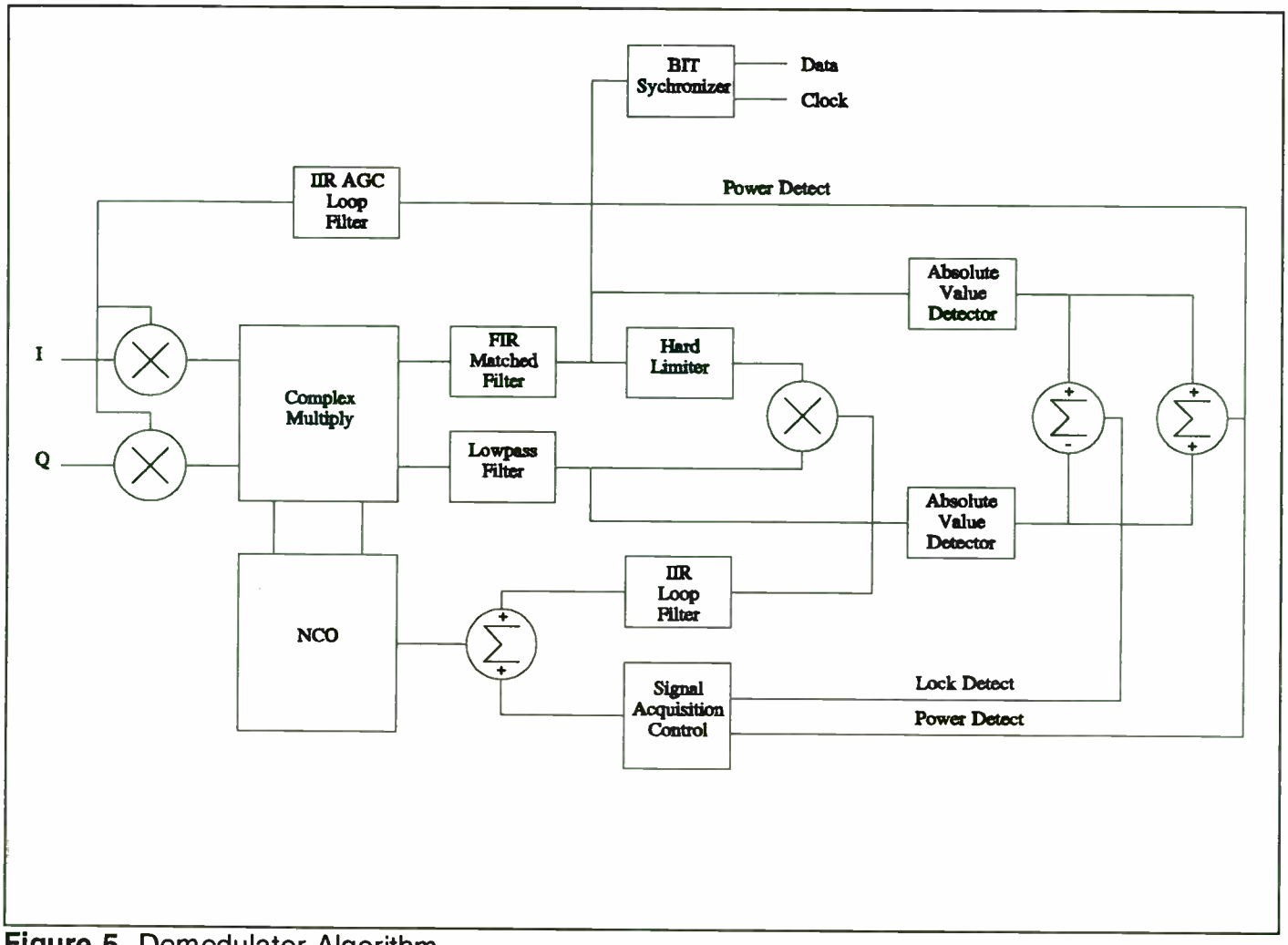


Figure 5 Demodulator Algorithm

implementation is speed. Analog circuits have the capability of operating at much higher data rates. As faster and faster digital components are developed, more and more analog systems will be replaced with digitized systems.

Figure 5 shows a block diagram of the demodulation algorithm. The algorithm is made up of three separate functions: 1) BPSK Carrier recovery loop (CRL) 2) Signal Acquisition Control for the CRL 3) clock regeneration using a bit synchronizer.

The BPSK CRL can be viewed as software implementation of a costas loop. The requirements that need to be met when a loop is designed in the analog world are the same as for loops designed in the digital world.

The digitized complex signal received from the RF front end is demodulated using the following software functions. Refer to figure 3.

Complex Multiplier - The input complex signal received from the RF front end is first multiplied with the complex signal received from the software NCO which is phased locked to the incoming signal. The resultant signals are the I channel (data channel for BPSK) and the Q channel (the tracking channel for BPSK).

FIR Matched filters - The I channel is filtered using the FIR matched filter discussed earlier. This FIR filter spans 8 data bits or is 32 samples long for signals that are sampled 4 times per bit. The Q channel is filtered using an integrator that has the same delay as the matched filter. Filtering of the Q or tracking channel is actually not required (data aided loop) but is done to make a less noisy lock

detect. An integrating filter is used instead of the FIR filter because the integrating filter uses less clock cycles than the FIR to implement.

Hard Limiter - The hard limiter outputs a '1' or '-1' depending on if the sample is positive or negative.

Multiplier - The multiplier simply takes two sample and does a floating point multiply.

IIR Loop Filter - Much work has already been done to analyze 2nd order type 2 loops in the analog world. This work can be used by designing the loop filter using an IIR filter. The weights for the IIR filter can be obtained by taking the transfer function of the filter and using the Bi-linear Z transform.

NCO - The software NCO (Numerically Controlled Oscillator) is basically the same as the hardware NCOs that are now available. The C40 processor keeps a record of phase using an integer 32 bits long. On each sample a delta phase is added to this integer. The first 8 bits of the integer are then used in a look up table that has the sine and cosine functions stored. Frequency is changed by changing the delta phase added to the integer.

The above software functions are all performed between adjacent samples. If the C40 operates at 40 MHz, the data rate is 50 kBs, and four samples are taken for each bit, all the above functions need to be done in 200 clock cycles or 100 instruction cycles (2 clock cycles = 1 instruction cycle). If higher speeds are desired multiple C40s operating in parallel could be used. Parallel processors can be added until the pipe line delays approach the delay of the loop filter.

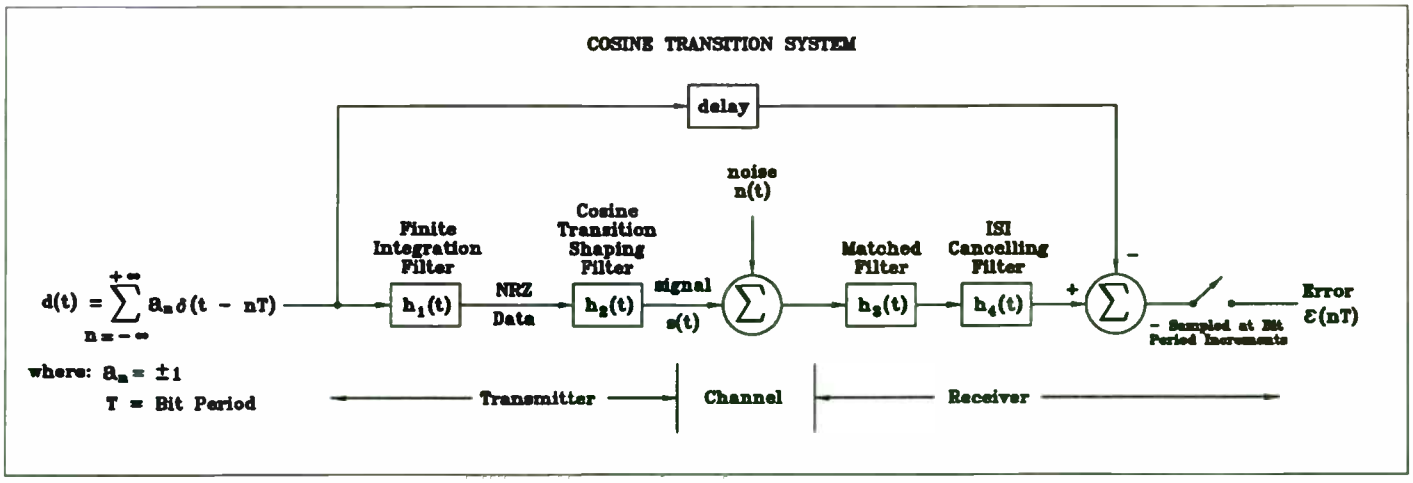
Signal Acquisition Control

For low data rate systems acquisition of the signal can be a problem. Low data rate systems have small loop bandwidths in comparison with the system's frequency uncertainty. As a result some scheme to acquire the signal must be implemented. One method is to sweep the CRL over the entire frequency uncertainty band and allow the CRL loop to lock when it finds the signal. This method can lead to false locks. A better method is to search for the maximum signal power over the band and pretune the CRL to that spot and then allow the CRL to lock. The maximum power can be found by tuning the LO over the sweep band and using an absolute value detector after the matched filter. Maximum power can also be found by using an FFT. A combination of the two techniques for finding maximum power is being implemented in UNISYs's demodulator.

Another important part of acquisition is knowing whether or not the CRL is locked or not. The lock detect indication used for the BPSK demodulator is FILTER(ABS(I)-ABS(Q)). This gives a good lock indication down to Eb/No of 0 dB.

BIT Synchronizer

The bit synchronizer is implemented using a conventional early late type phase detector. An IIR filter is used for the loop filter and the error signal generated is sent to the Sampling Clock located near the A/D converters. This loop determines where the samples are taken during the bit period which allows for best performance.



Receiver Filter Design

The transmitter filter is a combination of finite-integration filter and the filter that creates cosine transitions. The finite integration filter impulse response is:

$$h_1(t) = \begin{cases} 1, & 0 \le t < T \\ 0, & \textit{otherwise.} \end{cases}$$

The transition shaping filter impulse response is:

$$h_2(t) = \begin{cases} \frac{\pi}{2T}\sin\left(\pi\frac{t}{T}\right), & 0 \le t \le T \\ 0, & \textit{otherwise.} \end{cases}$$

The composite transmitter filter is the convolution of $h_1(t)$ with $h_2(t)$. The effective transmitter filter impulse response is:

$$h_1(t) \otimes h_2(t)$$

$$= \begin{cases} \frac{1}{2}\left[1-\cos\left(\pi\frac{t}{T}\right)\right], & 0 \le t \le 2T \\ 0, & \textit{otherwise} \end{cases}$$

$\otimes$ - *signifies convolution.*

The matched filter is, by definition, a time reversed and delayed version of the composite transmitter filter. Because the effective transmitter filter's impulse response is a real and even function, the matched filter is the same as the effective transmitter filter.

By convolving the matched filter with the transmitter filter, one can determine if any intersymbol interference (ISI) has been introduced.

$$h_3(t) = \left[h_1(-t) \otimes h_2(-t)\right]^*$$

$$= h_1(t) \otimes h_2(t)$$

$*$- *signifies complex conjugate*

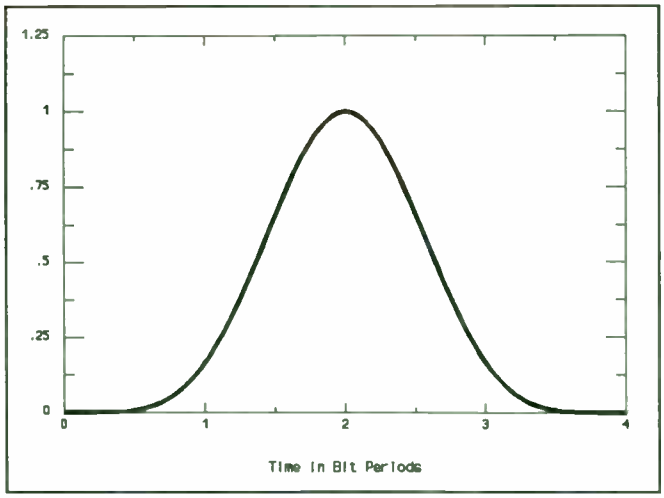


Figure 6 Impulse Response of Transmitter and Matched Filter

The calculated weights show that the matched filter will indeed introduce ISI. The weights

$$h_1(t) \otimes h_2(t) \otimes h_3(t) =$$

$$\begin{cases} \frac{t}{4} + \frac{t}{8}\cos\left(\pi\frac{t}{T}\right) - \frac{3T}{8\pi}\sin\left(\pi\frac{t}{T}\right), \\ \text{for } 0 \le t \le 2T, \\ T - \frac{t}{4} - \frac{t}{8}\cos\left(\pi\frac{t}{T}\right) + \frac{T}{2}\cos\left(\pi\frac{t}{T}\right) \\ + \frac{T}{8\pi}\sin\left(\pi\frac{t}{T}\right) + \frac{T}{4\pi}\sin\left[\pi\frac{(t-2T)}{T}\right], \\ \text{for } \quad 2T < t \le 4T, \\ 0, \text{ otherwise.} \end{cases}$$

correspond to the impulse response of the transmitter and matched filters at plus and minus one bit period on either side of the impulse delay of 2 bit periods (see figure 6). At time T and 3T the amplitude of the impulse response is T/8; at time 2T the amplitude is 3T/4. Figure 7 shows the resulting eye pattern with ISI.

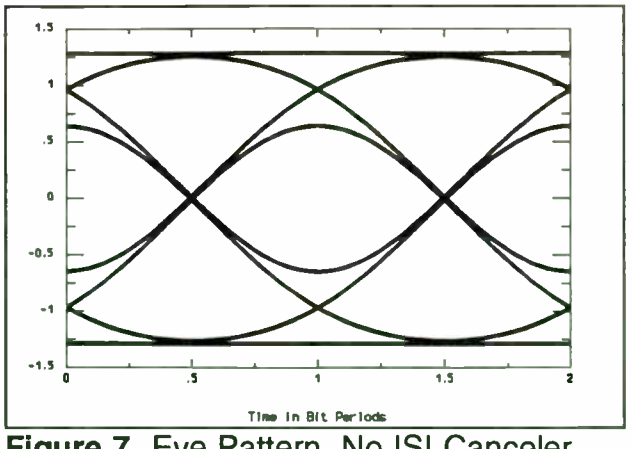


Figure 7 Eye Pattern, No ISI Canceler

Obviously, the ISI would have a severe impact on bit error rate. Therefore, an ISI canceling transversal filter is necessary and would need to have the following impulse response:

$$h_4(t) = \frac{4}{3T}\left[-\frac{T}{8}\delta(t) + \frac{3T}{4}\delta(t-T) - \frac{T}{8}\delta(t-2T)\right]$$

$$= -\frac{1}{6}\delta(t) + \delta(t-T) - \frac{1}{6}\delta(t-2T)$$

Since $h_4(t)$ is comprised of delta functions, convolutions involving $h_4(t)$ are easily calculated:

$$h_4(t) \otimes f(t) = -\frac{1}{6}\int_{-\infty}^{\infty} \delta(\tau)\, f(t-\tau)\, d\tau + \int_{-\infty}^{\infty} \delta(\tau - T)\, f(t-\tau)\, d\tau - \frac{1}{6}\int_{-\infty}^{\infty} \delta(\tau - 2T)\, f(t-\tau)\, d\tau$$

$$= -\frac{1}{6}f(t) + f(t-T) - \frac{1}{6}f(t-2T)$$

The resulting system impulse response when the ISI canceling filter, $h_4(t)$, is combined with the matched filter, $h_3(t)$, is shown in Figure 8.

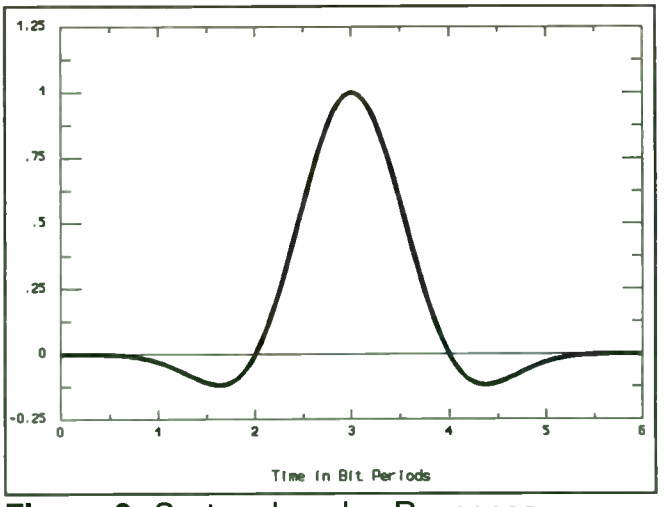


Figure 8 System Impulse Response (1st Iteration ISI canceler)

There still appears to be ISI. This is verified

by looking at the resulting eye-pattern, figure 9.

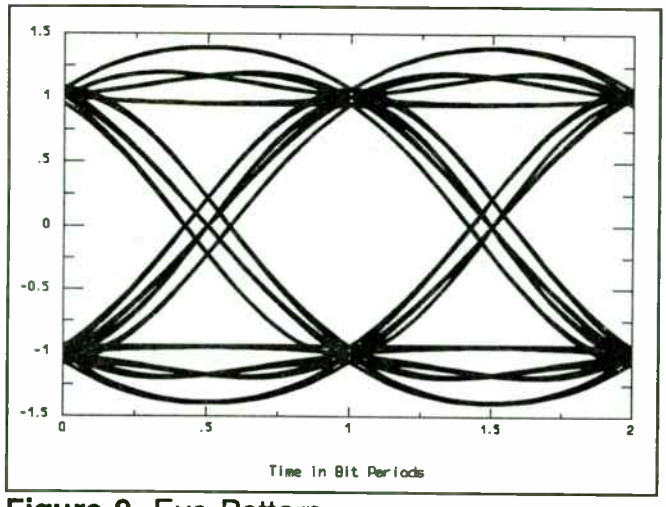


Figure 9 Eye Pattern
(1st Iteration ISI Canceler)

A modified ISI canceling filter is determined by subtracting by the correlation at 2 bit period offsets (see figure 8). The modified system impulse response is shown in figure 10:

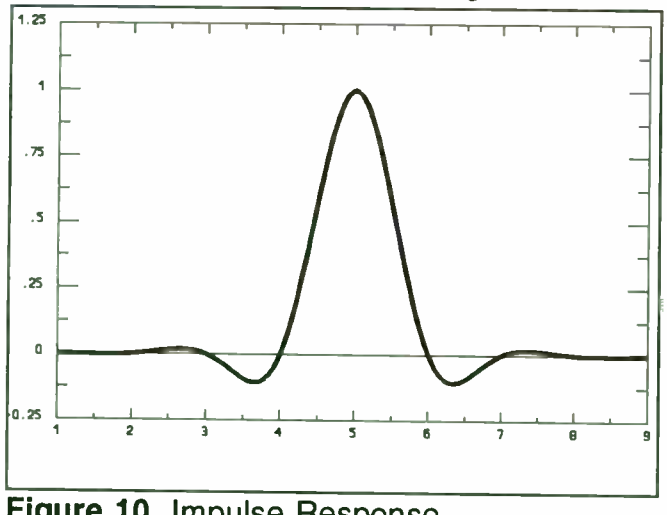


Figure 10 Impulse Response
(Modified ISI Canceler)

The correlation at bit period increments from the peak is now zero! The eye-pattern should have no ISI as shown in figure 11!

And, indeed, no ISI. The only question now to be answered is what kind degradation from ideal is the bit error rate when this system is used.

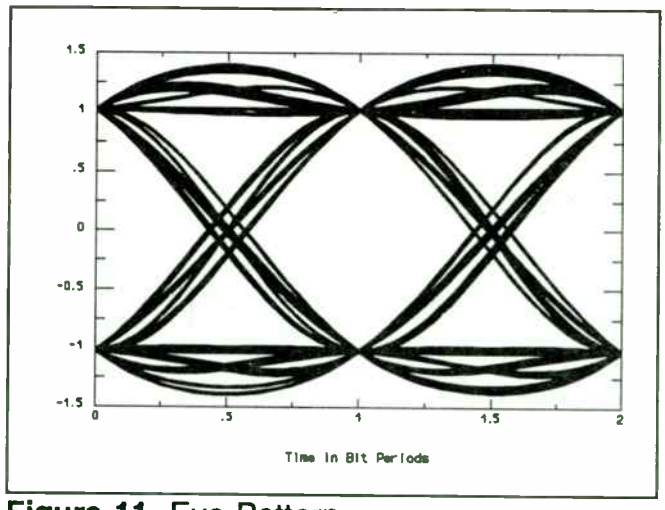


Figure 11 Eye Pattern
(Modified ISI Canceler)

Since the actual filters were implemented with digital FIR algorithms, the number of taps required and tap spacing needs to be determined. First the receiver impulse response is determined.

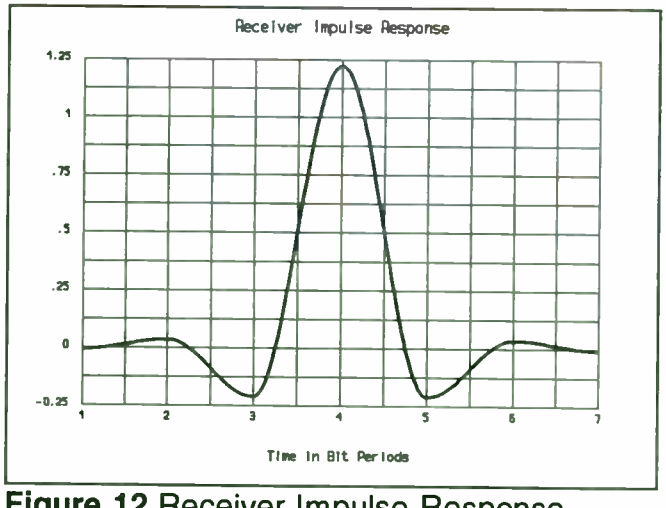


Figure 12 Receiver Impulse Response

It is advantageous to use the minimum number of taps as possible in the FIR filter. The greater the number of taps, the longer the processing time required, and the lower the data rate allowable. We analyzed a variety of spacing possibilities and ended with a sampling rate of 4 times our highest data rate. This was convenient because it allowed us to implement an early/late gate bit synchronizer with 1/2 symbol spacing. Also, by comparing

the equivalent noise variance of an ideal finite integration filter (matched filter for non-bandlimited NRZ data) with our receiver filter, we analytically determined the implementation loss to be 0.25 dB from ideal! It later was verified through simulation.

Designing a High Performance Monolithic PSK Modulator

Frank Hayden and Robert J. Zavrel
GEC Plessey Semiconductors
1500 Green Hills Road
Scotts Valley, CA 95066

The objective of this program was to build a low power, monolithic, high performance PSK modulator. There are a wide variety of applications for such a device. The primary objectives was to fabricate a direct sequence "spreader" that could operate up to 1 GHz and provide excellent carrier suppression and combine the spreading PRN code with the data for both CDMA cellular and Part 15 applications. A second objective was to provide a high performance PSK modulator for more traditional applications, including TDMA and point-to-point microwave systems. A third objective was to facilitate derivative products that can provide QPSK, OQPSK and MSK baseband signals.

The "SL801" has been fabricated and tested. Proceeding into a true product development depends upon market acceptance of the device. Therefore, this paper represents a report on a development project and should not be considered a "product" introduction.

1. Digital Section

The SL801 is a Bi-Phase shift keyed (BPSK) modulator. The device accepts data in a digital format along with a spreading code that is used to modulate an RF carrier for transmission in spread spectrum systems. A primary performance goal was to maximize carrier suppression. The chosen circuit topology reflects this primary design objective in that two modulators are cascaded thus the carrier supression characteristics are greatly enhanced.

Referring to the block diagram of figure 1, asynchronous data is fed into a type D flip flop at the data input. This serves to synchronize the data to the code clock. Note that the code clock is clocking all the flip flops and latches. After clock synchronization of the data, code inversion modulation is performed by modulo-2 addition of the code and the synchronized data. This takes place in exclusive OR gate G1. This serves to embed the code and data and also adds a greater level of security to the transmitted information. All the user must remember to do in the receiver is to modulo-2 add the received baseband information with the same spreading code in the receiver. This approach ensures that good auto correlation and low cross correlation are preserved.

Referring to the block diagram, Figure 1 , at point A we have

$$(1) \quad A = C \oplus D$$

Where:
A is the output of exclusive OR gate G1.
C is the Spreading Code.
D is the synchronized data.

At point A the output of G1 is fed into another D type flip flop that serves to delay the code and data by one full code clock cycle. The output of the flip flop at point B may be expressed as;

$$(2) \quad B = (C \oplus D)\tau$$

Where:

B is the output of the delay flip flop L2.
τ is the clock delay through L2.

At this point in the circuit, point B is exclusive OR'ed with point A giving output C which can be expressed as;

$$(3) \quad C = (C \oplus D) \oplus (C \oplus D)\tau$$

Where:

C is the output of exclusive OR gate G2.

In order to assure proper cancellation of the undesirable terms and thus maximizing carrier suppression, the modulation inputs to the two multipliers must be correctly phased. Latches L3 and L4 assure that the modulation input signals are fed simultaneously to MOD2 and MOD1 respectively.

2. Modulator Section

The carrier input to MOD 1 and MOD 2 require a differential drive. The unbalanced input is converted to the required differential input by the carrier buffer, a unity-gain amplifier-balun with a maximum allowable input signal level of -5 dBm.The balanced modulators (MOD 1 and MOD 2) are Gilbert Cell multipliers configured such that the operation of the circuit provides balanced bi-phase modulation. Each modulator has 0 dB gain to minimize the effect of offsets that could corrupt the balance of the modulator and hence compromise carrier suppression. The modulators are cascaded so that the carrier suppression of each modulator will add to effectively double the carrier suppression (dB).

The instantaneous phase of the RF carrier is determined by the instantaneous polarity of the modulating signal. Consequently, two phase states, 0 or π radians (BPSK) are permitted. In this respect the

balanced BPSK modulator serves as an analog exclusive OR gate. The output of the balanced modulator can be expressed as follows;

(4) $V_{out} = \omega_c \oplus V_{data}$

Where: V_{out} is the output voltage of the modulator. ω_c is the RF carrier signal to be modulated. V_{data} is the modulating data signal.

To facilitate maximum carrier supression, unity gain in the modulator path is used. Unity gain eliminates amplification of the offset voltages. Offset in the modulators leads directly to a decrease in the carrier suppression in the modulator.

The carrier buffer is simply a differential amplifier that takes a single ended input and gives a differential output. The inputs are biased at approximately 3 volts and the input impedance is approximately 10 k ohms and 2 pF. If a 50 ohm source is to be used to drive the SL801, simply terminate the input with 50 ohms and capacitivly couple the signal to the input. The other differential input is AC grounded through 50 ohms in series with 10 pF. The 10 pF capacitor is connected to a pin on the package that allows the low frequency response of the input to be extended by adding a capacitor from this pin to ground. Emitter degeneration is used in the amplifier to extend the linear signal handling ability of the amplifier up to a maximum of -5 dBm.

The Gilbert cell modulator schematic is shown in Figure 2. The schematic shows the basic Gilbert cell and the level shifters at the modulating inputs and the carrier inputs. this level shifting ensures that the transistors in the cell do not saturate, which could degrade carrier suppression.

The lower two transistor pairs Q33 and Q3 are the RF input and the four upper transistors Q26, Q13 and Q4 ,Q24 are the modulating or phase reversing transistors. Emitter degeneration is used in the lower pair of transistors to extend the linear signal handling capability of the RF input. The upper four transistors in the cell have no degeneration as this would degrade the gain of the modulator.

As the upper pairs of transistors are switched the RF carrier is steered to the modulator output depending on which set of transistors in the top of the cell are biased on. For instance if Q 26 and 24 are on, then RF is passed to the output through the cascoded pair made up of Q33 and Q26 and Q3 and Q24. As Q13 and Q4 are turned on by the modulating wave form the cascode arrangement switches over to Q33 and Q13 and also Q3 and Q4. So as the modulating wave form changes state the phase of the carrier is changed 180 degrees. The gain of the modulator circuit can be approximated by the following equation;

(5) $G_v = \dfrac{RI}{(R_e + r_e)}$

Where;

G_v is the voltage gain of the circuit.

RI is the collector load resistance. (R38 or R28 in the schematic)

r_e is the intrinsic emitter resistance of the transistor where $r_e = \dfrac{IC}{V_t}$.

Since the modulator path has 0 dB of gain , both of the modulators are identical. Now returning to the block diagram in Figure 1, the output of the first modulator can be expressed as follows with appropriately substituting (3) into (4);

(6) $D = \omega_c \oplus \left[(C \oplus D) \oplus (C \oplus D)\tau\right]$

where ω_c is the RF carrier signal, and

(7) $E = \omega_c \oplus \left[(C \oplus D) \oplus (C \oplus D)\tau\right] \oplus \left[(C \oplus D)\tau\right]$

Now using the following Boolean identity;

(8) $A \oplus B \oplus B = A$

(9) $E = \omega_c \oplus (C \oplus D)$

which is the desired result for the output of MOD 2.

So by using the delay component, two modulators may be cascaded together to achieve approximately twice the carrier suppression that can be achieved with one modulator. It should also be noted that the output from the first modulator is indeed a spread spectrum signal. However, there is no way that the receiver can demodulate the information that is in this signal unless the appropriate delay is generated in the receiver.

It is also worth noting that the choice of the delay component is quite critical to the success of the cascaded arrangement of modulators. It is important that the two modulating signals be as closely decorrellated as possible. Therefore, the minimum clock delay for correct operation is one, and the larger the delay, the more decorrellated the two signals become, but at the expense of added circuitry. It was felt that the minimum delay period would yield a satisfactory result.

3. Driver Section

After modulating the RF carrier, the output of the modulators is driven into a differential output 50 ohm driver. The output of this differential driver can be configured in several ways. Since the output is an

open collector circuit, a 50 ohm resistor connected to each output and up to the supply rail will properly set the gain of the driver and allow the circuit to drive up to -5 dBm into any 50 ohm load. The output signal can also be driven into a balun and an additional 6 dB of drive obtained, less any balun losses.

The 50 ohm driver circuits are open collector differential amplifiers. A bias resistor of 50 ohms must be connected to the open collector outputs up to the 5 volt supply line. These loads can then be AC coupled to the next circuit in a single ended fashion, coupling one of the outputs to the next stage and also AC coupling the unwanted output to a 50 ohm load. This terminates the collectors with the correct differential impedance. Either output can supply -5 dBm, or if a RF balun is used correctly, the output power would increase to +1 dBm.

4. AGC Amplifier

The modulated signal is filtered externally before entering the input to the AGC amplifier. Alternatively, further processing and/or phasing with other SI801s can yield QPSK signals. The AGC amplifier has 5 dB of gain and 30 dB of AGC control range.

The AGC amplifier is also a Gilbert cell type circuit. Referring to the schematic of the amplifier in figure 3, the main portion of the amplifier is made up of transistors Q13,Q81,Q11,Q19,Q4 and Q54. In parallel with this Gilbert cell is another amplifier made up of transistors Q79,Q80,Q60 and Q33. This amplifier in parallel with the Gilbert cell serves to keep the DC point at the load resistors R55 and R53 constant as the gain control is varied by diverting current away from the collectors of Q11,Q19,Q4 and Q5.

The gain of the AGC amplifier is controlled via the differential amplifier comprised of transistors Q15 Q56. as the voltage at the AGC control input is increased above the DC bias point at the base of Q15, Q56 turns on and the current through the diode load Q57 supplies the bases of Q19 and Q4. The diode current compensates for the nonlinear control characteristic and provides a linear control characteristic over the voltage range of 2.5 to 3.5 volts.

5. Output Amplifier

The output of the AGC amplifier is fed into a differential amplifier capable of driving 0 dBm into a 50 ohm load. This differential amplifier has 0 dB gain and is similar to the 50 ohm driver at the output of the modulator section. The outputs of the amplifier are pulled up to the supply via 50 ohm resistors and capacitvely coupled to the load. If a differential drive is not desired the unused output should be capacitivly coupled to a 50 ohm load.

This 50 ohm driver is similar to the 50 ohm driver that is at the output of the modulators, only capable of supplying 0 dBm to the load in a single ended application or again if an RF balun is used it is capable of delivering +3 dBm.

6. Design Considerations

On the chip are separate voltage regulators and grounds for the digital and analog functions. The chip can be put in a standby mode by supplying a TTL compatible signal that is active high.

In the digital section of the device all TTL/CMOS input levels are level translated down to a peak swing of 150 mV on chip. This is done to avoid large signal swings at the modulating inputs of the Gilbert cell and also helps with cross talk on chip.

Separate bandgap regulators for the digital and analog functions on the chip are used to set the bias point for the current sources . This was done to reduce the cross talk among the various digital and analog circuits.

In the layout of the I.C., great care was taken to match all of the transistors and metal runs in both the digital and analog sections. In the modulator section of the device all metal traces were matched to within 1 milli-ohm to avoid DC offsets. The carrier buffer, two modulators and the first 50 ohm driver are also surrounded with trench isolated substrate contacts to further avoid cross talk from the digital section. Both the AGC amplifier and the AGC 50 ohm driver are also surrounded with trench isolated substrate contacts to ensure that the carrier does not leak into these circuits degrading the carrier suppression after the modulator section.

7. The HE Process

Process HE is a bipolar process incorporating polysilicon base/emitter contacts, trench isolation and an advanced base/emitter structure. The process has been characterized as having a peak Ft of 14 Ghz. There are three layers of metalization available on the process.

The NPN devices are available in four different emitter lengths of 2, 5, 10 and15 microns. There are also PNP devices available. Internitride capacitors can also be fabricated on the process with realizable values of up to 5 pF for a single capacitor. Inductors can be fabricated on the third metal layer with Q's measured at 10. There are also two different types of resistors, both of which are polysilicon. Low value resistors can be realized using 200 ohms per square polysilicon and high value resistors can be made by putting a barrier around the 200 ohms per square polysilicon increasing the sheet resistivity to 2000 ohms per square. Another feature of process HE is the ability to open selective substrate contacts around different portions of the circuitry and then surrounding these substrate contacts with walled trench isolation. This then serves to isolate portions of the circuit from other functions on the chip, thereby increasing the amount of isolation between different circuit functions.

8. Conclusions and Further Work

The objectives of designing a high performance, low power, and convenient PSK modulator have been met. Figure 5 shows a typical output spectrum of the device.

Two SL801 circuits can be combined to produce QPSK or OQPSK signals. Furthermore, MSK can be generated with two SL801s operating in quadrature. MSK is the most likely baseband of choice for frequency hopping spread spectrum systems operating under Part 15. These applications are certainly possible using the currect SL801 and/or derivative circuit configurations.

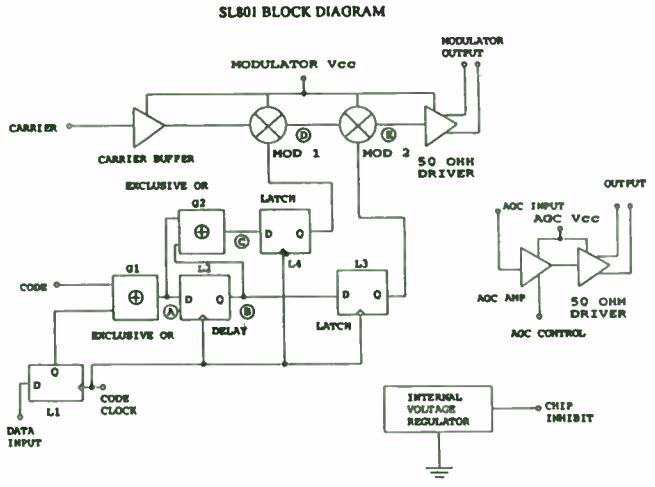


Figure 1.

MODULATOR

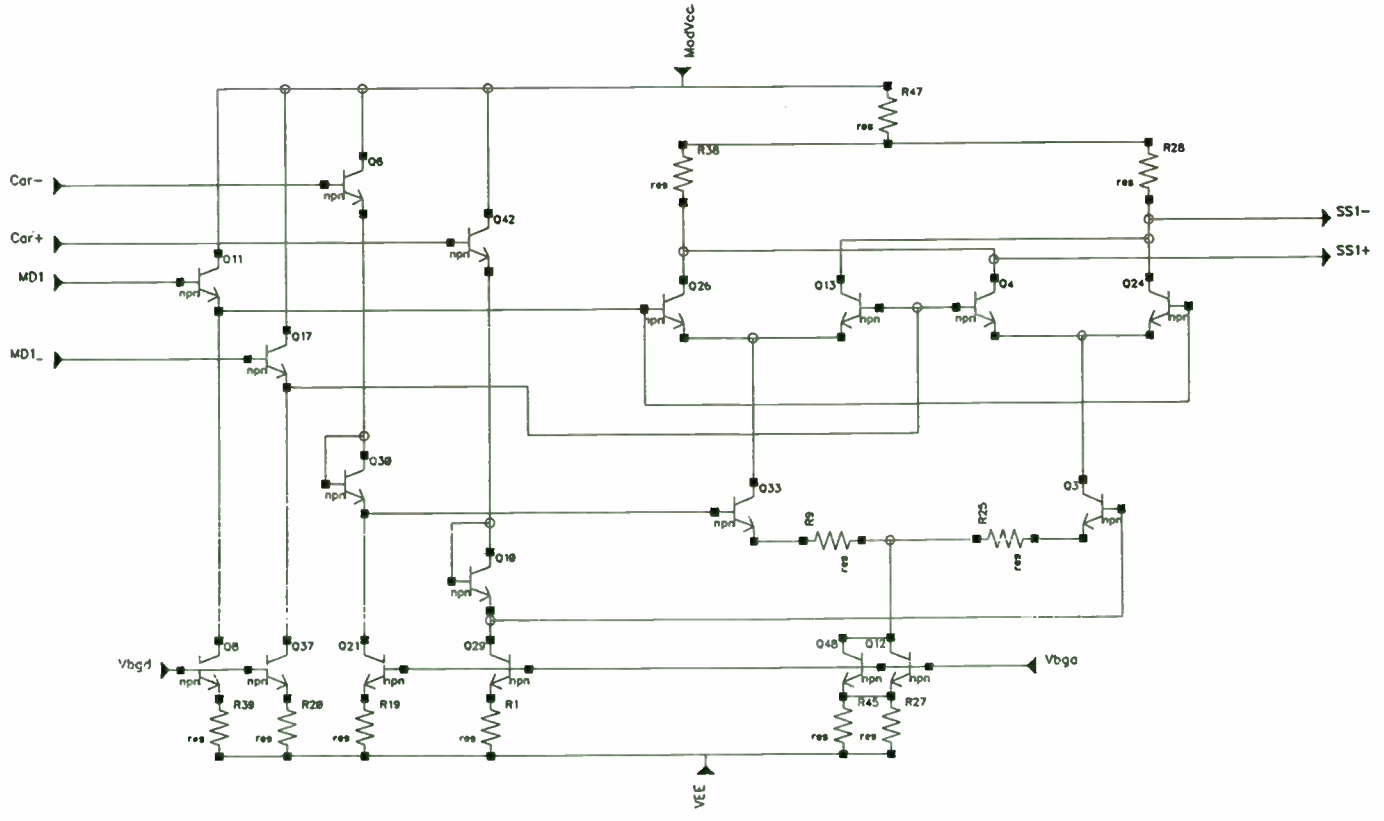


Figure 2.

AGC AMP

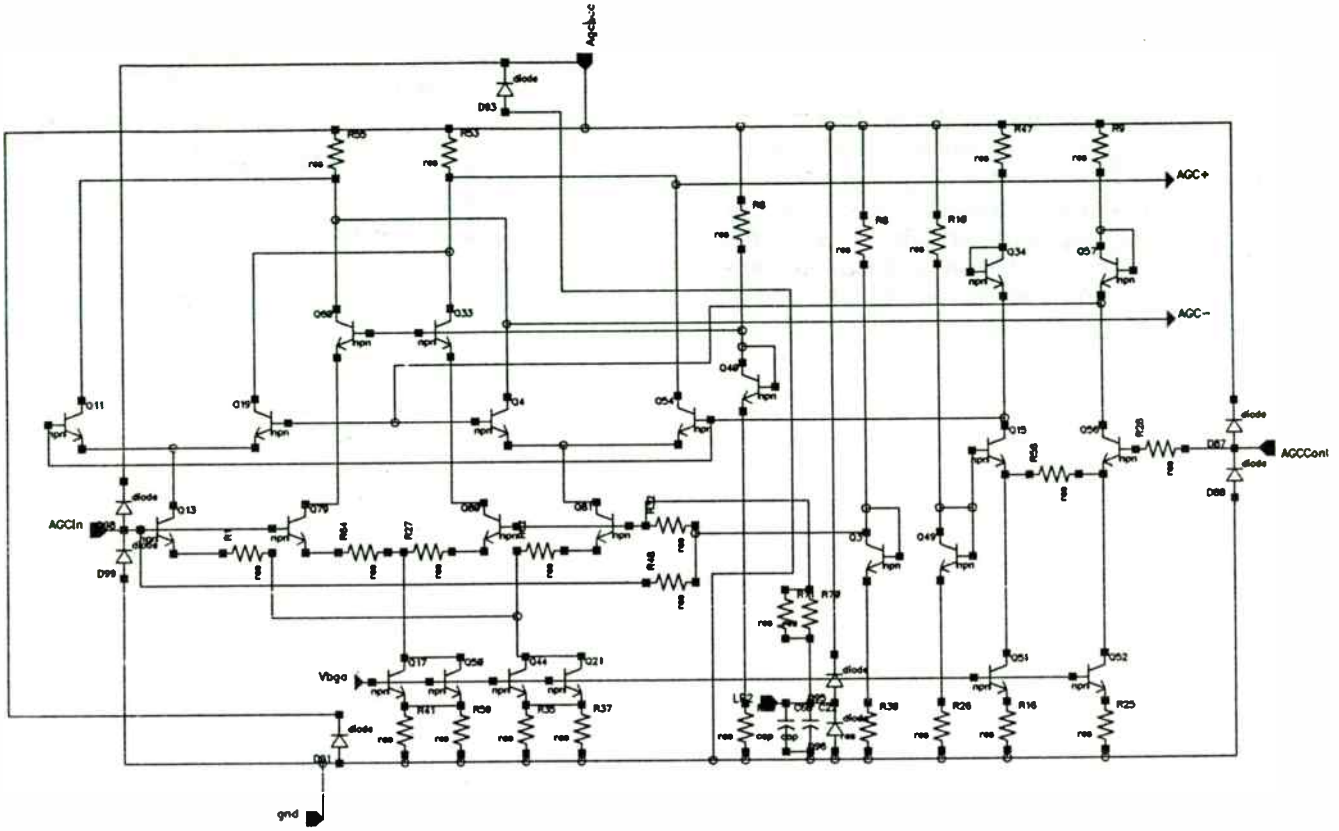


Figure 3.

50 Ohm DRIVER

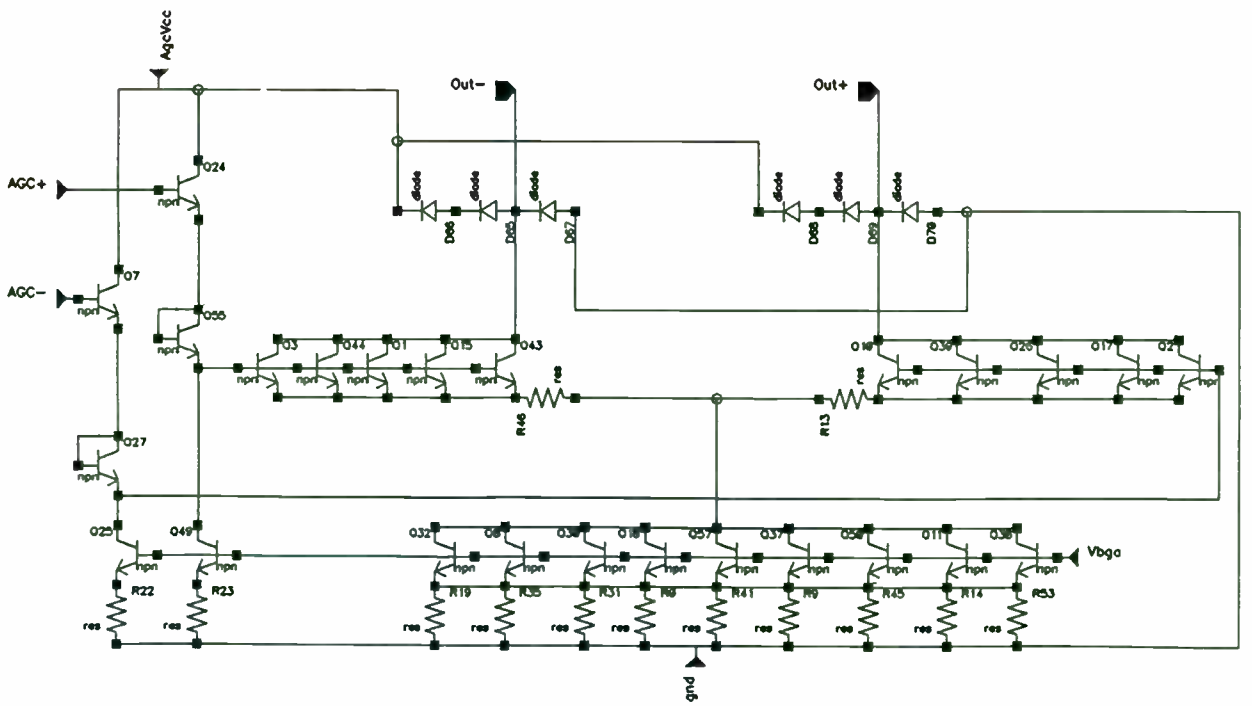


Figure 4.

TYPICAL OUTPUT SPECTRUM

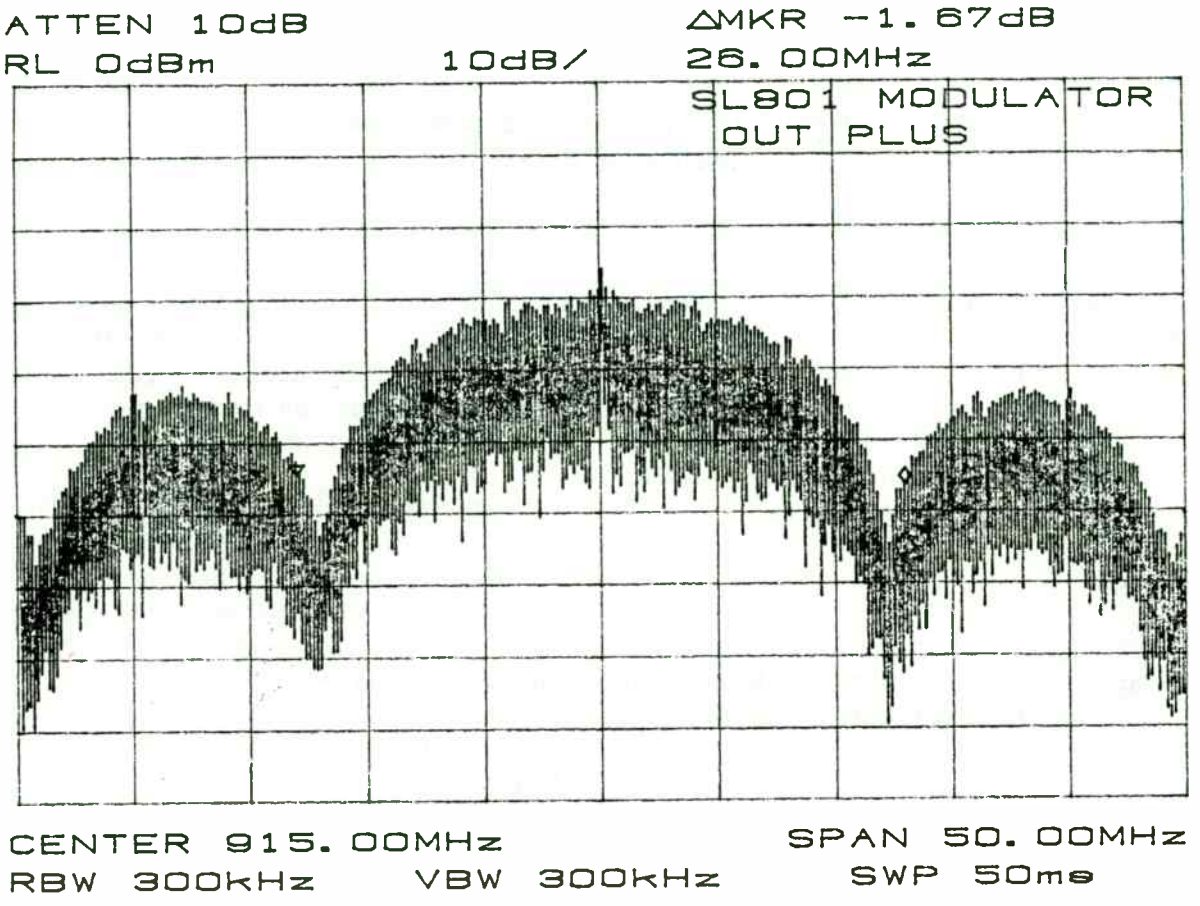


Figure 5.

Methods for Estimating and Simulating the Third Order Intercept Point

Carl Stuebing and Mojy C. Chian

Harris GCSD and Harris Semiconductor
Melbourne, FL 32901

Abstract - The third order intercept point (P_{ip3}) is a metric to quantify the nonlinearity of a circuit. P_{ip3} is used to describe the adjacent frequency isolation and gain distortion. This paper describes the concept of P_{ip3} and reviews methods to estimate and simulate it. The common simulation pitfalls are identified and practical solutions are discussed and illustrated.

I. Introduction

The nonlinearity of a radio amplifier determines it's distortion. As the input signal level increases, distortion of the output signal becomes prominent and undesired harmonic products are produced.

Consider two FM carriers at 50MHz and 100MHz applied to the input of a radio amplifier. Assume that there is little filtering of the input signal and that we are interested in amplifying the 100MHz signal. However, the 50MHz signal is so strong that it drives the amplifier into distortion. In this situation, the second (2X50MHz) harmonic of the 50MHz signal would interfere with the 100MHz signal at the output of the amplifier.

Obviously, radio receivers generally do have selective filters on their signal inputs. So interference of signals is, as a practical matter, more important for signals which are close together in frequency (i.e. in band).

Nonlinearity can also cause less obvious interference problems with signals that are close together in frequency. When two in band signals are applied to a nonlinear circuit element, they will mix. This process is called intermodulation and the resultant frequency components are referred to as intermodulation products. It can be shown that the frequencies of the intermodulation products are defined by $k_1 f1 \pm k_2 f_2$, where k_1 and k_2 are integers. The component with a frequency equal to $n_1 f1 \pm n_2 f_2$ is defined as the nth order intermodulation product, where $n = n_1 + n_2$ [1,2].

While the two fundamental signals will typically be magnified by the amplifier, components at the sum and difference frequencies are also produced at the output. These second order intermodulation products are not usually considered to be important, because they are typically far enough out of band from the fundamentals that they can be easily suppressed with filters.

More important are the third order intermodulation products. These occur at frequencies given by the two times one frequency minus the other, and vice versa. While these parasitic signals may be smaller in amplitude than the second order products, they are more difficult to filter out. Consider two FM signals at 100 and 101 MHz. The third order parasitic tones are produced at 202-100=102MHz and 200-101=99MHz. When the signal distortion is large, these products become significant since they are produced well within the band of the input frequencies.

A popular metric to quantify distortion is the Third Order Intercept Point, or "P_{ip3}". For a two-port network excited with two sinusoidal signals with frequencies f_1 and f_2, if the third-order intermodulation product output power ($P_{2f_1-f_2}$) and the output power at f_1 (P_{f_1}) are plotted versus the input power at f_1, the third-order intercept point (P_{ip3}) is defined as the point where P_{f_1} and $P_{2f_1-f_2}$ intercept. The third-order intercept point is a theoretical level, however, it is a useful and popular quantity to estimate the third-order intermodulation products at different power levels (Figure 1).

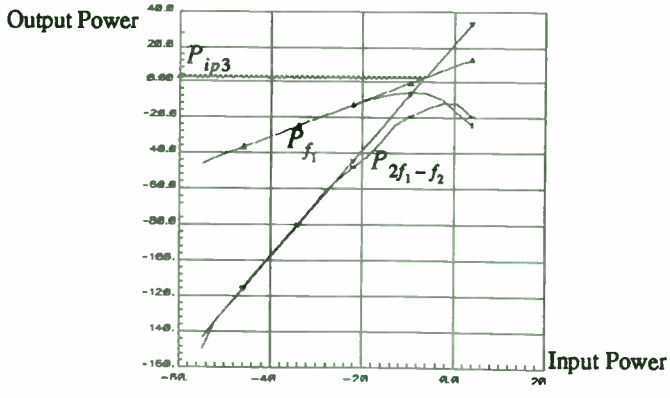


FIGURE 1. P_{f_1}, $P_{2f_1-f_2}$, and their asymptotes.

The power gain of the fundamental frequencies is linear and remains constant until we have a large enough input signal that clipping or gain compression occurs. However, the voltage gain of the third order products is raised to a cubed power instead of linear (as in the case of the fundamental). This means that the power gain of the third order harmonics is three times higher than that of the linear fundamental. Of course, the power level of the third order product is normally less than that of the fundamental. As the input power level increases, the ratio of the fundamental power output to the third order product output power level decreases rapidly.

The output P_{ip3} (P_{ip3o}) is defined as the point where the asymptotes of the fundamental linear power gain curve and the third order product power gain curve intersect. We can use simple trigonometry to relate this output power level to the fundamental and third order frequency gain curves. Knowing P_{ip3o}, the input power level, and the power gain, we can extrapolate the power level of the third order intermodulation products. If G is the power gain, then

$$P_{ip3o} = P_{ip3i} + G$$

When P_{ip3} is large, the power level of the third order products is suppressed, or distortion is reduced. Obviously, this is desirable. When P_{ip3} is small, the third order product power is relatively large, or distortion is increased.

While the "real world" does not consist of only two interfering radio channels, this is a convenient technique. Here we allow ourselves to define a metric for quantifying the distortion and thus the interference of radio signals.

II. Review of Some Traditional Estimation Techniques

Several "rules of thumb" exist within the RF industry for estimating P_{ip3}. Since accurate simulation of P_{ip3} can be very computer time intensive, one should attempt first to estimate this parameter. P_{ip3} is in some ways more dependent on system level considerations (such as the input termination method and available power supply voltage and current) than device design considerations (such as the geometry selected for the RF transistor). Generally, once the system determines P_{ip3} (which should be quickly albeit roughly estimated), computer simulation can be used to measure the somewhat subtle effects of transistor or integrated circuit level design on P_{ip3}.

Rule #1

The first (system level) rule of thumb is that P_{ip3} is about 6 to 10dB higher than the 1dB gain compression power level. The 1dB gain compression level refers to the signal level at which the gain at the frequency of interest has decreased due to clipping by 1dB or 12%.

There are no known elegant mathematical discussions to prove this rule. We simply have an empirical number which has been known to be a useful approximation. This is not an extremely accurate rule, but coupled with a one tone simulation is much faster than two or three tone simulations.

Rule #2

The second rule of thumb relates output power available to output intercept point. As an approximation, use the maximum (RMS) power available to the load as the 1dB compression point for output power.

This rule implies that the amplifier is piecewise linear. Thus, gain is approximated to always be equal to the (very) small signal level until the available output power level has been reached. Once the peak output power has been exceeded, the signal is suddenly clipped and incremental gain becomes zero at the maximum or minimum voltages of the output. The point at which this process starts is used to identify the approximate one dB compression point.

Real amplifiers typically demonstrate increasing compression with amplitude. Thus, 12% gain compression normally occurs below the maximum available output power level. Therefore, this rule of thumb is very rough and optimistic.

Extension #1

Some interesting results can be obtained from carrying the above assumption further. Assuming that voltage saturation of devices is not permitted, the available output power for a single collector load resistor in a common emitter amplifier is given by:

$$P_{out1dB} = \frac{I_{nom}^2 R_l}{2}$$

Where I_{nom} is the nominal DC bias current in the collector of the transistor. This is assumed to be the available zero-to-peak output current. Of course, this equation neglects the lost power output (i.e. gain) at higher frequencies due to capacitive loading.

Rule #2 can now allow us to take this available output power and use it for 1dB gain compression estimates. First, we rewrite the above equation in dBm:

$$P_{out1dB} = 10\log(500I_{nom}^2 R_l)$$

Now, we apply Rule #1 to estimate Output Intercept:

$$P_{ip3o} = 10\log(500I_{nom}^2 R_l) + 6dB$$

Extension #2

A popular method for obtaining Input Intercept is to find the difference between P_{ip3o} and power gain. However, a shortcut can simplify the calculation of P_{ip3o}. Power gain depends on input impedance, output impedance, and current gain (i.e. β). This information can also be used to determine the available output power and input power levels that are associated with the estimated 1dB compressed output power. No explicit gain estimate is really necessary. This method, generally, works best at frequencies much less than the f_T of the transistor.

Thus at the 1dB compression point, output signal power can be related to input signal power. The RMS input signal power is given by:

$$P_{in} = \frac{I_b^2 R_{in}}{2}$$

where I_b is the nominal base bias current associated with I_{nom}. We observe that the input current "just below compression" (given our simplified, piecewise linear model) is simply specified by I_{nom} divided by β (which is a function of frequency). The input resistance is given by r_π and R_b:

$$P_{in1dB} = \frac{I_{nom}^2 (r_\pi + R_b)}{2\beta^2}$$

Since $r_\pi = \frac{\beta V_t}{I_{nom}}$, thus [3]:

$$P_{in1dB} = \frac{R_b I_{nom}^2}{2\beta^2} + I_{nom}^2 \left(\frac{\beta V_t}{2\beta^2 I_{nom}}\right)$$

therefore,

$$P_{in1dB} = \frac{R_b I_{nom}^2}{2\beta^2} + \frac{I_{nom} V_t}{2\beta}$$

Since we are considering frequencies much lower than the f_T of the device (where β is large and R_b should be much less than r_π), the above equation can be approximated by:

$$P_{in1dB} = \frac{I_{nom} V_t}{2\beta}$$

To convert the 1dB output compression level to P_{ip3o}, apply rule #1, and add 6 to 10 dB.

$$P_{ip3i} = 10\log\left(\frac{500 I_{nom} V_t}{\beta}\right) + 6dB$$

What is particularly interesting about this result is that P_{ip3i} (in Watts, not dBm) is proportional to V_t, I_{nom} (nominal collector current) and $1/\beta$. Of all the above terms, the only process related one is β. The rest are dependent on absolute temperature and the available power supply current. So system constraints dominate this situation. Also notice that large β is "bad", because it reduces P_{ip3i}. Of course, large β tends to help noise figure.

This rough method is much faster than lengthy computer simulations, and can give us a reasonably good feel for the system level trade offs affecting Pipi3.

III. Simulation tools and techniques for Computing P_{ip3}

The traditional method for calculating P_{ip3} by simulation is to use a Harmonic Balance based simulator. Harmonic Balance is a frequency domain method applied to nonlinear circuits, where the computations are performed using the trigonometric-series coefficients. The approach is based on balancing of currents between the linear and nonlinear subcircuits in the frequency domain. Since nonlinear devices are generally expressed in the time domain (e.g. Gummel Poon model for a BJT) their response has to be determined using time domain techniques. First, the input signals to the nonlinear devices are converted to time domain signals (using IFFT). Second, the time domain signals (voltages) are applied to the nonlinear devices and the time domain response (currents) is determined. Third, the time domain currents are converted to the frequency domain (using FFT). Finally, the frequency domain currents of the nonlinear devices are used in a global circuit equation to satisfy KCL at each node. This process is repeated many times until a consistent solution is attained (convergence).

Harmonic Balance determines the steady state response of nonlinear circuits to sinusoidal excitations. Even

though the frequency domain response can be converted to the time domain, it does not contain the transient response of the circuit. This feature is a mixed blessing. For circuits with a very long and unwanted transient response (e.g. switching power supplies), harmonic balance is an efficient technique to bypass the transients and obtain the steady state response. For such cases, time domain based analysis algorithms can waste a lot of CPU time analyzing an undesired region of the response. On the other hand, for circuits like LNAs and comparators, it is essential to observe the transient response of the circuit to determine its effect on the frequency domain response. This issue is discussed in more detail in Section IV.

Time domain simulators (e.g. SPICE [4]), solve the global circuit equations by solving the nodal analysis matrix equation, $GV = I$, to determine node voltages (V). They solve an $n \times n$ system of equations per time step, where n is the number of nodes. Harmonic Balance simulators, on the other hand, satisfy the nodal current equations (KCL) for all harmonics in one step. This means that the system of equations to solve are $(nm) \times (nm)$. Where m is the desired number of harmonics. Since the CPU time required to solve a system of equations superlinearly increases with the size of the matrix, Harmonic Balance is expected to perform well for small circuits but not for large circuits. The large size of the system of equations also creates huge memory requirements even for moderately sized circuits by RF standards (e.g 10 transistors). As an example, a two tone RF circuit with 10 transistors may require as much as 600 MegaBytes of disk space. Additionally, since the convergence difficulty very rapidly increases with the number of equations involved, it is expected (and observed) that Harmonic Balance will have significantly more convergence problems than SPICE.

SPICE type simulators are very efficient in performing time domain analysis on large circuits. Transient analysis of circuits with many thousands of transistors is routinely performed by IC designers. In fact, RF ASICs (typically with less than 100 transistors) are considered small by IC simulation standards.

The drawback on using SPICE appears when the results are converted to the frequency domain. The typical process of performing transient analysis and then using FFT normally does not provide enough accuracy in the frequency domain. The following simple example demonstrates the lack of sufficient actuary (also referred to as the dynamic range of FFT). Consider the following example of two series sinusoidal sources. A typical SPICE simulation, with 5005 internal time points, followed by an FFT, with 1024 points, offers only 80db of dynamic range, Figure 2.

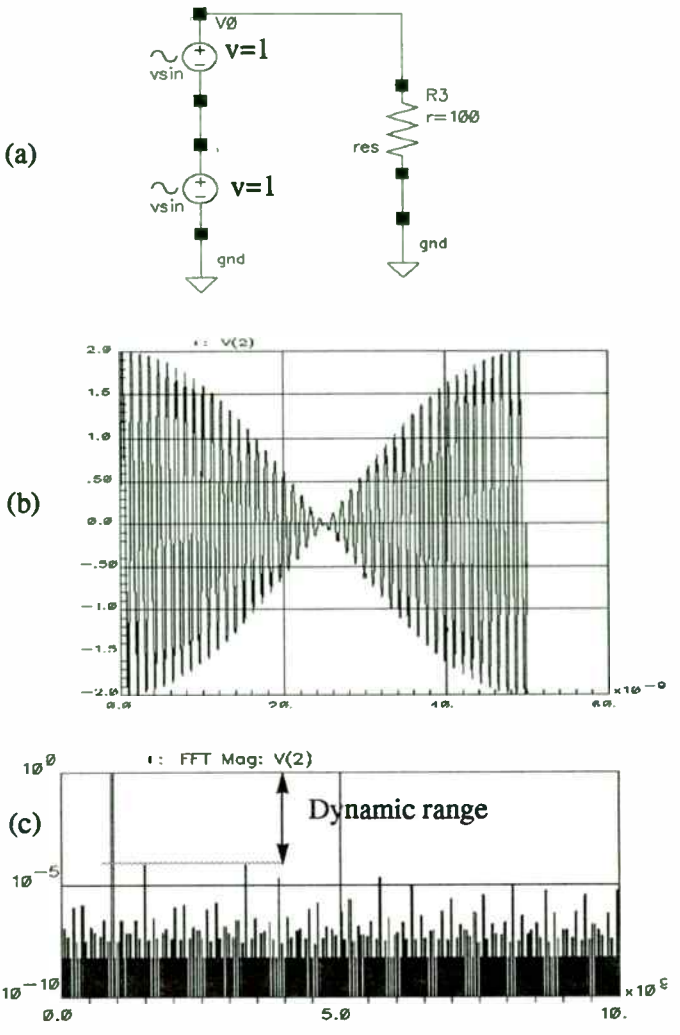


FIGURE 2. Two series sinusoidal sources (a), the time domain waveform (b), FFT results (c) with 5005 internal simulation time steps and 1024 FFT sampling points

In Section V, we will investigate the typical low dynamic range FFT and propose a technique to remedy the problem. With such a technique, the dynamic range for the same circuit and the same or less number of time points can be increased to 260 dB or more.

IV. Non Steady State and Other Pitfalls

RF circuits are traditionally designed, simulated and built with narrow band topologies. The narrow band approach is generally desirable and necessary. However, narrow band configurations, as well as time constants inherent in amplifiers, can cause design and simulation problems.

For example, consider a perfectly linear amplifier with unlimited bandwidth and no input capacitance. If we connect the amplifier to a two tone source which is AC cou-

pled, we will have a simple and convenient example for simulating the impact of nonsinusiodal steady state on the dynamic range of a system (Figure 4a).

If we choose a 50 Ω source, 50 Ω load and 1000pF for AC coupling, we have produced a 100nsec time constant ($\tau = rc$). Suppose that the fundamental frequencies are at 100MHz and 110MHz. The beat frequency would be 10MHz, having a period of 100nsec (intentionally chosen to be equal to the time constant).

When we start a two tone simulation in the time domain, we are really driving the circuit with a unit step multiplied by each of the two tones. Thus, we are not able to start the simulation at time=0 in sinusoidal steady state.

This observation is consistent with the behavior of coupling capacitors. At t=0+, the voltage across the coupling capacitor remains zero, such that no apparent phase shift is observed initially. Once sinusoidal steady state has occurred, the phase shift of the two tones across the coupling capacitor has been established and remains stable. The response of the system can be expressed as

$$V_r(t) = Ke^{-\frac{t}{rc}}u(t) +$$

$$(A\sin(\omega_1 t + \phi_1) + B\sin(\omega_2 t + \phi_2))\,u(t))$$

Let us assign $A = B = 1$ for convenience. This forces K to become a relative number. The impedance of the capacitor at 100MHz and 110MHz is given below.

$$|Z_{c1}| = \frac{1}{2\pi f_1 C} = 1.59\Omega$$

$$|Z_{c2}| = \frac{1}{2\pi f_2 C} = 1.45\Omega$$

We can then calculate the steady state phase angles.

$$\phi_1 = \text{atan}\left(\frac{|Z_{c1}|}{R_1 + R_2}\right) = 0.91°$$

$$\phi_2 = \text{atan}\left(\frac{|Z_{c2}|}{R_1 + R_2}\right) = 0.83°$$

K is determined by using the steady state phase angles at t=0+.

$$V(0) = 0 = Ke^0 + \sin(0 + 0.91°) + \sin(0 + 0.83°)$$

$$K = 30mV/V = -30.5dB$$

The first four beats of the voltage drop across the capacitor in the circuit of Figure 4a in response to 1V amplitude tones is shown in Figure 3.

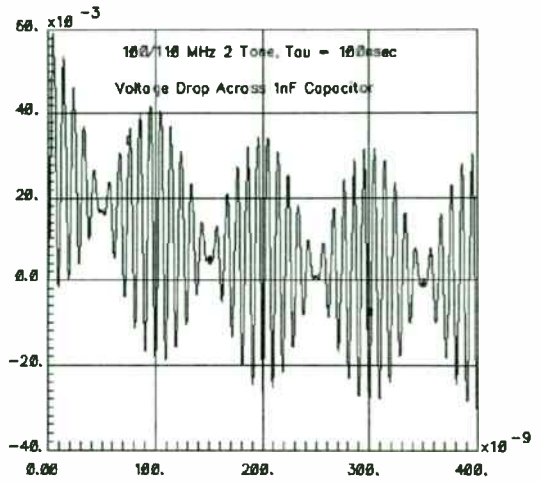


FIGURE 3. The first 4 beats of the time domain voltage across the capacitor in Figure 4a.

Notice how we have no voltage drop at t=0+, and that the envelope has a 100nsec time constant, with an initial value of about 30mV, as calculated.

If the capacitor voltage envelope is allowed to settle one time constant before we perform an FFT of the time domain results, the time domain instability decreases by e. This is down to 36.8% or -8.69 dB of initial value.

Suppose that we run a two tone time domain simulation for two beats, or 200nsec using the circuit in Figure 4a. Also, suppose that we allow the circuit 100nsec (one time constant) to settle, and attempt to obtain sinusoidal steady state. Then we perform an FFT of the results from 100nsec to 200nsec. At 100nsec, or one time constant, the impact of the transient response has decreased to 0.368K or 1.1% of the signal amplitude. But at 200nsec, the ending time of the FFT, the transient has decreased to 0.135K or 0.4% of the signal amplitude. Thus, we have 0.368K-0.135K = 0.7% of instability in our simulation. This would lead us to expect to see no worse than 20log(0.011-0.004) = -43dB of dynamic range in our simulation due to the transient effect.

Figure 4a shows the frequency spectrum if we simply wait one time constant for settling (which might be considered a easy error to make). Of course, the transient in the time domain produces broadband spurs, so that no single frequency has the entire 43dB of error. We obtain only 60dB of dynamic range in this simple circuit. This would be considered unacceptable by most designers. Figures 4c and 4d indicate the tremendous improvement in dynamic range if we wait to 5 or 25 time constants

before doing the FFT. Of course, the ultimate solution to this problem is to eliminate the capacitor entirely with either a wire or a battery.

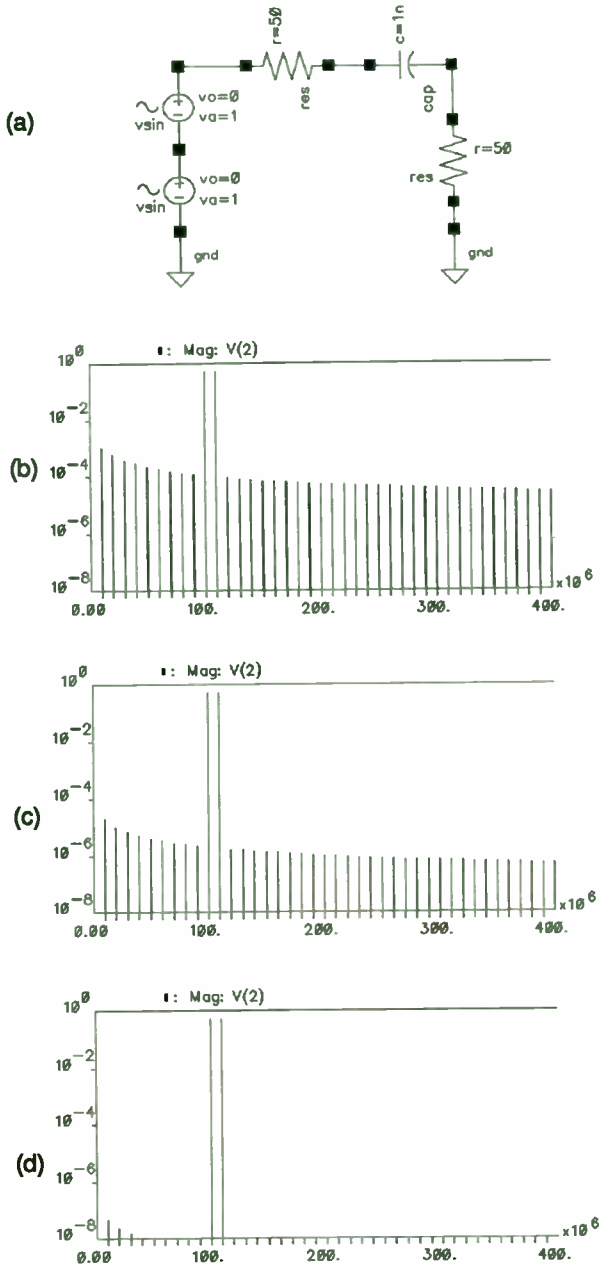


FIGURE 4. A simple two-tone circuit (a), the frequency spectrum across the load resistor, where the FFT samplings were started after 1τ (b), 10τ (c), 100τ (d).

A simple alternative to a battery or a short is a larger capacitor. The trick here is that the capacitor needs to be much larger, such that $rc \gg T_{FFT}$. This approach may seem entirely the opposite of what we have just advocated. But the real goal is to minimize transient effects. This is accomplished by either choosing the starting time of the FFT to be much larger than the circuit time constants or forcing the circuit time constants to be much larger than the period of the FFT (beat period).

The improvements resulting from using bigger capacitors are shown in Figure 5. Making the capacitor considerably larger (1000X and 10000X respectively) can dramatically reduce spurs related to transients in the time domain.

Another way to avoid these problems is to use a Harmonic Balance simulation. Harmonic Balance can be particularly effective because it will not have "noise" spurs in the results due to transients in the time domain.

The danger of only using Harmonic Balance techniques is that time domain transients, as a result of signal modulation, may produce undesirable spurs in the real circuit that are not simulated. A design may seem noisy, because the coupling circuitry interferes with the modulation of the signal. As a result, the expected system dynamic range may not be obtained.

Consider a simple case of the implicit unit step function (as discussed previously) being replaced by a pulse train. While this pulse train would be typically much slower than the signal which it is modulating, a problem might be lurking. If this modulation pulse train is close to $1/\tau$ for some important system time constant, this signal would become distorted. The spurs that we saw in the earlier example would be observed as repeating and causing the signal to appear to have poor dynamic range.

We do not recommend finishing an RF design without a check with a time domain simulator like SPICE to verify

that transients in the time domain are not causing problems.

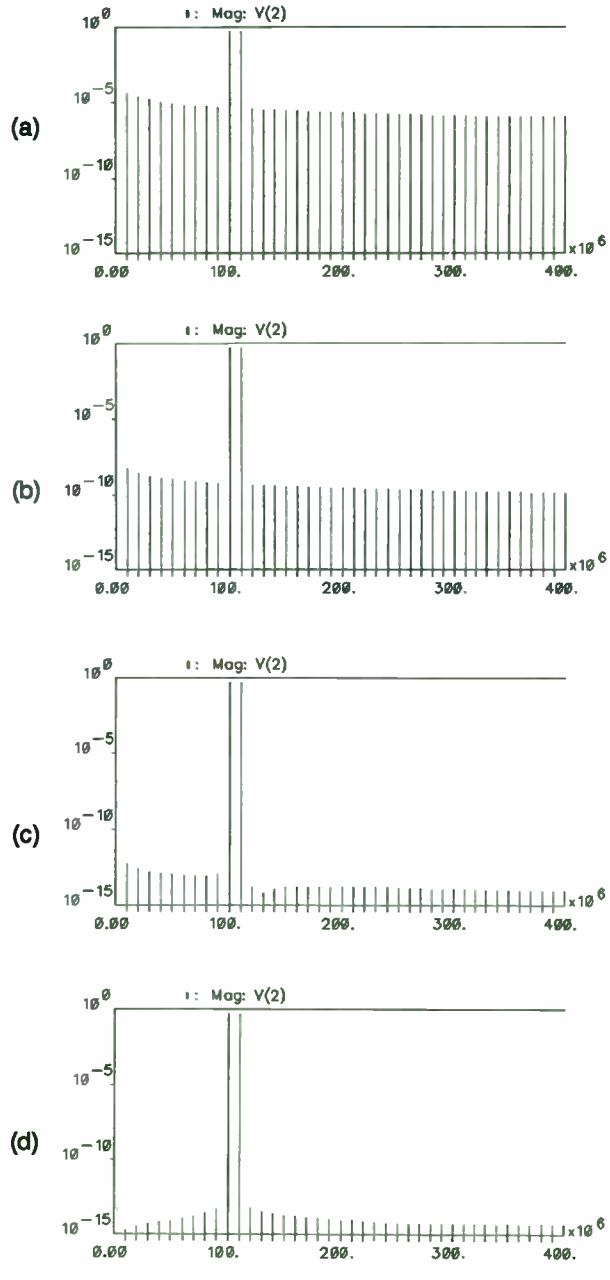


FIGURE 5. The spectrum of the voltage across the load resistor in the circuit shown in Figure 4a with the coupling capacitor set to 10n(a), 1u (b), 100u (c), and shorted (d).

V. Methods for Accurate and Efficient Simulation of P_{ip3} Using SPICE

Even when steady state is reached, the typical method of SPICE transient analysis followed by an FFT of the results to observe the nonlinearity in the frequency domain is an inaccurate (noisy) method. The major source of inaccuracy (low dynamic range) stems from the interpolation error. Since FFT is a post processing of the transient results, generally, the simulation time steps and the FFT sampling steps are not synchronized. An interpolation between two adjacent simulation points is required to determine the value of the time domain waveform at an FFT sampling point (Figure 4).

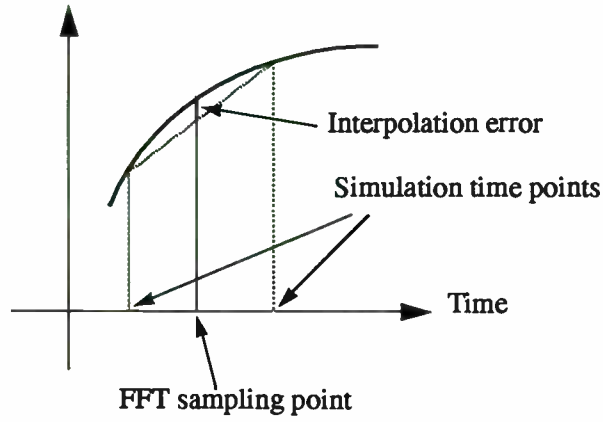


FIGURE 6. FFT Interpolation Error

Unfortunately, in typical SPICE type simulators, the user can not force the internal time points to land on specific locations.

In order to eliminate the interpolation error, the FFT sampling steps need to be predicted before the transient simulation is performed. The simulator, then, must be forced to step onto the sampling points. In the large signal AC analysis section of FIRST (Fastrack's Integrated RF Simulation Tools [5]), the following algorithm is used to eliminate the interpolation error.

1- Determine the frequency of the input sources.

2- Determine the Beat frequency, f_b. It can be shown that the beat frequency is the largest common divisor of the input frequencies.

3- Determine the starting point in time for the FFT sampling, t_{init}. This is the point (user specified) at which we assume that the circuit has reached satisfactory steady state.

4- Determine the end point for the FFT sampling, t_{end}. Where

$$t_{end} = t_{init} + \frac{1}{f_b}$$

5- Determine the largest non-negligible frequency content of the desired signal, f_{max}. This is the frequency at which the spectrum of the desired signal effectively dies out. Even though exact knowledge of this frequency is not required, it plays an important role in eliminating aliasing.

6- Determine the number of FFT sampling points, N_{fft}. Where

$$N_{fft} \geq \frac{2f_{max}}{f_b} \quad , \qquad N_{fft} = 2^n \text{ , and } \quad n \blacksquare integer$$

See Appendix A for proof.

7- Determine the FFT sampling points, t_i. Where

$$t_i = t_{init} + \frac{i}{f_b N_{fft}} \qquad i = 0, \ldots, (N_{fft} - 1)$$

8- Run the simulator and force it to step onto the FFT sampling points.

9- Perform an FFT of the time domain results from t_{init} to t_{end} using N_{fft} sampling points.

As an example, the above method was used on the circuit shown in Figure 4a. The simulator took 1029 time steps. The frequency spectrum is shown in Figure 7.

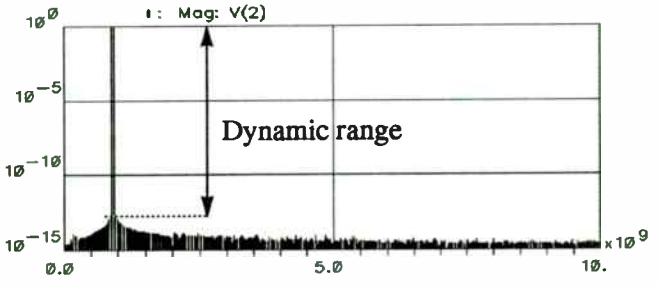


FIGURE 7. FFT results with Interpolation error eliminated

Even though the internal time points for the simulation were approximately 5 times less than the original attempt, the dynamic range of the FFT jumped from 80db to over 260dB

As another example, an LNA (4 transistors) with 900 Meg and 920 Meg input signals was simulated with FIRST to determine P_{ip3}. Figure 8 shows the frequency spectrum of the output with a 7 mV amplitude for the input signals. Next, the input signal levels were varied to generate the output power plots. Figure 9 shows the power plots with and without the asymptotes.

From Figure 9, we can infer that Pip3o is about 2dBm. Notice how the output fundamental begins to compress at about -8dB (down 10dB as expected). Even the third order product's power curve is a straight line over 30dB of input signal dynamic range. This demonstrates sufficient dynamic range in the simulation to get consistent and useful results.

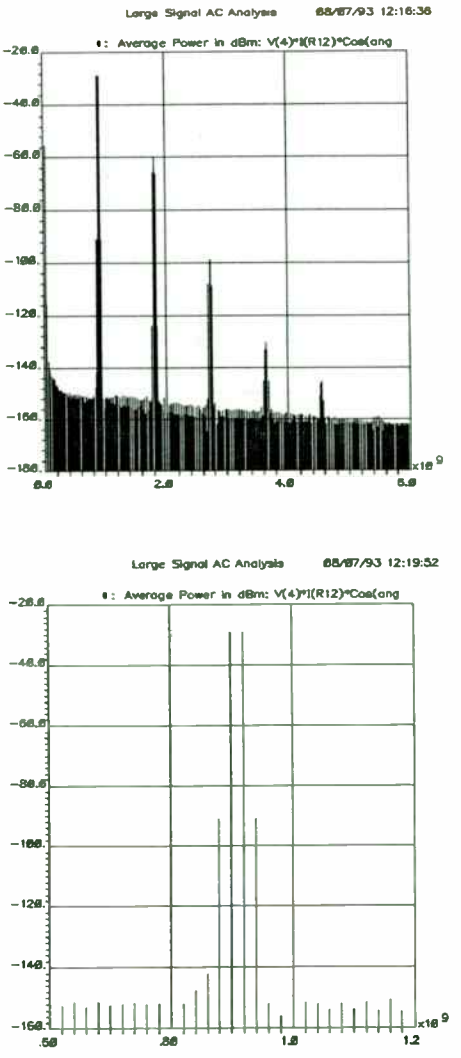


FIGURE 8. The entire frequency spectrum (a) and the spectrum around the fundamental frequencies

V. Practical Example of Estimation, Simulation, and Actual Measurements on a Mixer

A single balanced mixer was designed and fabricated with Harris Semiconductor's UHF-1 process. UHF-1 is an 8 GHz, silicon bipolar integrated circuit process. The power budget allowed 5mA for the bias current, and 5

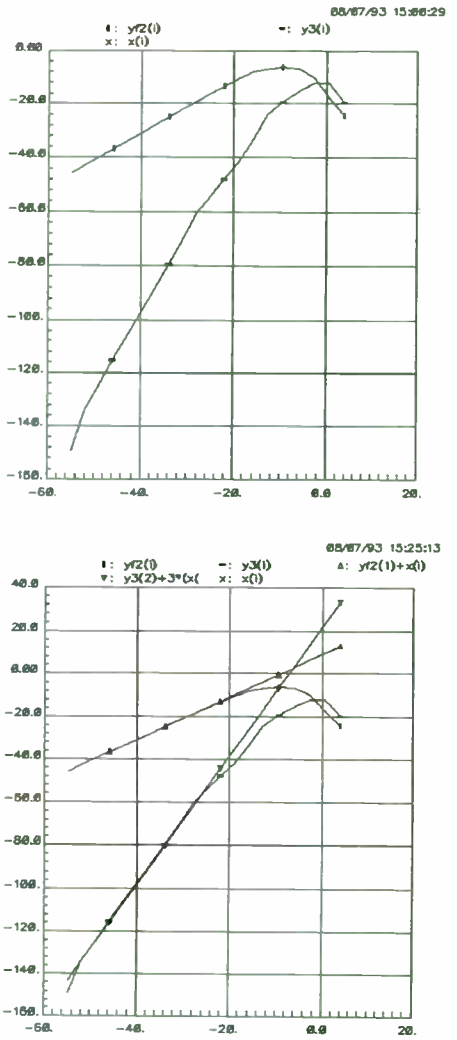


FIGURE 9. The power plots for the LNA (a) and the power plots with the superimposed asymptotes

Volts for the power supply. The part is operated in a 50 Ω environment at about 900MHz.

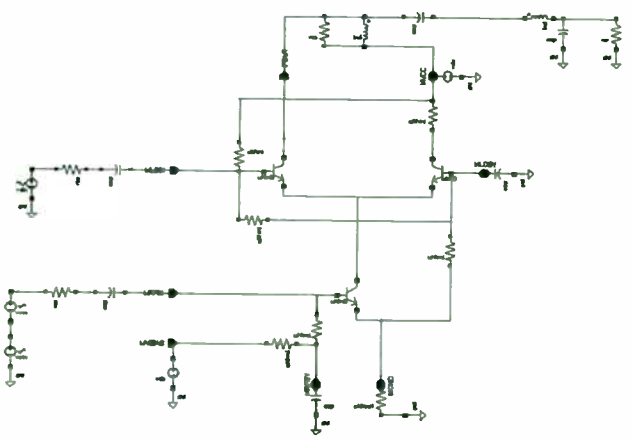


FIGURE 10. A simplified schematic of a UHF-1 single balanced mixer.

Several steps are required to estimate the P_{ip3o} of this mixer. Notice that the nominal collector current of the common emitter RF device within the mixer is 5mA. Two resistors are connected to the LO collector. One serves as a source and the other as a load, such that both are in parallel. However, the external load is a 50 Ω resistor which is reactively transformed to appear as 1K Ω to the mixer. The mixer output source resistor is 1K Ω (Figure 10).

The nominal zero to peak current allowed to the external load appears like 2.5mA into 1K Ω (as described in extension #1). However, the single side band mixing process reduces this by a factor of $\sqrt{2}$, leaving 1.77mA available.

$$P_{avail} = \frac{I_{nom}^2 R_L}{2} = \frac{(1.77mA)^2 (1K\Omega)}{2} = 1.57mW$$

$$P_{avail} = 2.0dBm$$

$$P_{ip3o} = 8 \rightarrow 12dBm$$

Experimental results indicate P_{ip3o} is +5 to +7dBm (depending on the package). Since conversion gain is +8dB, P_{ip3i} is -3 to -1dBm. Simulation results using FIRST for 40mv, 50mV, and 63mv signal amplitudes are tabulated below. Notice that the estimate is optimistic, while the simulation is more accurate.

Amplitude	40mV	50mV	63mV
P_{f1}	-20.45	-18.50	-16.50
$P_{2f_2-f_1}$	-76.31	-71.18	-65.10
P_{ip3o}	7.48	7.84	7.80

VI. Conclusions

P_{ip3} is a popular metric used by RF designers to specify the third order intermodulation products. Several rules of thumb, based on system level considerations are used to approximate P_{ip3}. However, simulation can be more accurate, and give a clearer indication of how subtle changes in a circuit impact P_{ip3}. The decaying transient response of a system and the interpolation error can cause the FFT results to appear noisy (picket fence effect). Neglecting these decaying transients (as in Harmonic Balance) may cause inaccuracy in predicting the dynamic range of the circuit. A methodology was presented to eliminate the interpolation error and thus significantly improve the dynamic range of the FFT. It was shown that by proper treatment of decaying time constants and the

interpolation error, P_{ip3} can be reliably measured by SPICE based time domain simulation followed by an FFT of the results.

VII. References

[1] G. Gonzalez, Microwave Transistor Analysis and Design, Printice-Hall, Inc., New Jersy, 1984.

[2] R. S. Carson, High Frequency Amplifiers, Second Edition, John Wiley & Sons, New York, 1982.

[3] P. R. Gray and R. G. Meyer, Analysis and Design of Analog Integrated Circuits, Second Edition, John Wiley & Sons, New York, 1984.

[4] L. W. Nagel, "SPICE2, A Computer Program to Simulate Semiconductor Circuits", Technical Report ERL-M520, UC Berkeley, May 1975.

[5] M. C. Chian, S. S. Majors, and A. G. Whittaker, "Embedding RF Design Tools in an IC Design System", the proceedings of RF Expo East, 1993.

Low Cost Phase Noise Measurement Technique

Jim H. Walworth
Tampa Microwave Lab, Inc.
12160 Race Track Road
Tampa, Florida 33626-3111
(813) 855-2251

Phase noise is one of the most important specifications of an oscillator and often one of the most difficult to measure. Unfortunately it can also be one of the most expensive test set configurations an oscillator manufacturer must invest in. Many times the only equipment a small company may have is a spectrum analyzer. Usually the phase noise of a good phase locked oscillator (PLO) will be lower than the measurement capability of the spectrum analyzer. This paper suggest several techniques of measuring phase noise, not necessarily covered previously in the literature.

This paper will offer four basic methods of measuring phase noise all with different price tags.

1. Fully automated phase noise test set such as the HP3047.
2. Measurement with the HP11729 Carrier Noise Test Set.
3. User modified (used) HP11729B test set Phase detector method
3a. Discriminator Method
4. Utilizing a PLL itself to measure it's phase noise with only a low cost low frequency analyzer.

The relative price tags of the above approaches are listed below. They are approximate. Number 3 has the greatest amount of uncertainty in the price since it assumes the availability of used equipment in the same order of magnitude found by the author.

RELATIVE PRICES OF THE FOUR TECHNIQUES.

1. HP3047A SYSTEM	> $110K
2. HP11729C + HP8662A	> $71.5K
3. HP11729B + SR770	< $13K
3a. HP11729C + SR770	< $13K
4. SR770	< $ 8K
(SR550)	< $ 6K

Method 1. Fully automated phase noise test set. (Figure 1)

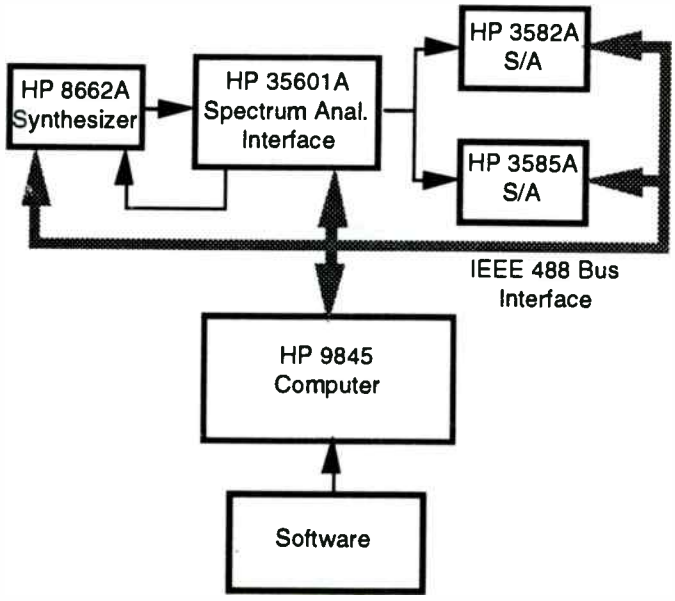


Figure 1. HP 3047 System [1], [2]

The largest advantage to this technique is the automatic calibration and compensation. All instrumentation is connected via IEE488 buss to collect error correction data, and measurement data. This is by far the best phase noise measurement system. It is also the most expensive.

Method 2. Manually calibrated and operated phase noise test set. Figure 2. shows the block diagram of a manually calibrated / operated test set up, consisting of the HP11729C Carrier Noise test set, the

HP8662A synthesizer, a low frequency spectrum analyzer, and a microwave spectrum analyzer.

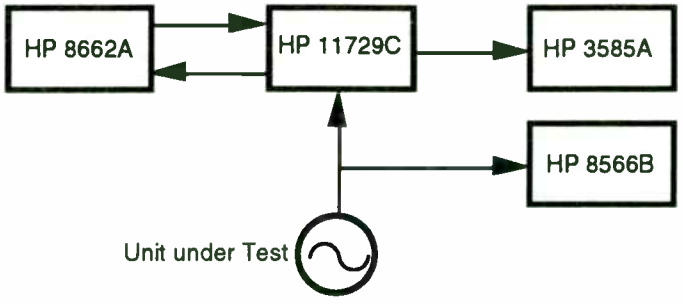


Figure 2. HP 11729C System

The block diagram of the HP11729B carrier level test set is shown in Figure 3. This test set has the basic components to make measurements in either the phase detector method or the discriminator technique. Both techniques are described in the HP Product Note 11729B-1

Method 3. HP 11729B phase noise test set and a low frequency spectrum analyzer.

This system can be used to measure phase noise without the synthesizer for specialized applications. The diagram in Figure 4 shows the HP11729B with one slight modification to allow the injection of a microwave VCO. (instead of the synthesizer) to mix with the unit under test (U.U.T.). The modification is not a difficult one since it only requires the removal of connectors inside the unit and the addition of another connector to the front or rear panel. We chose to remove the "AUX" connector on the front panel and replace it with an SMA panel mount bulkhead.

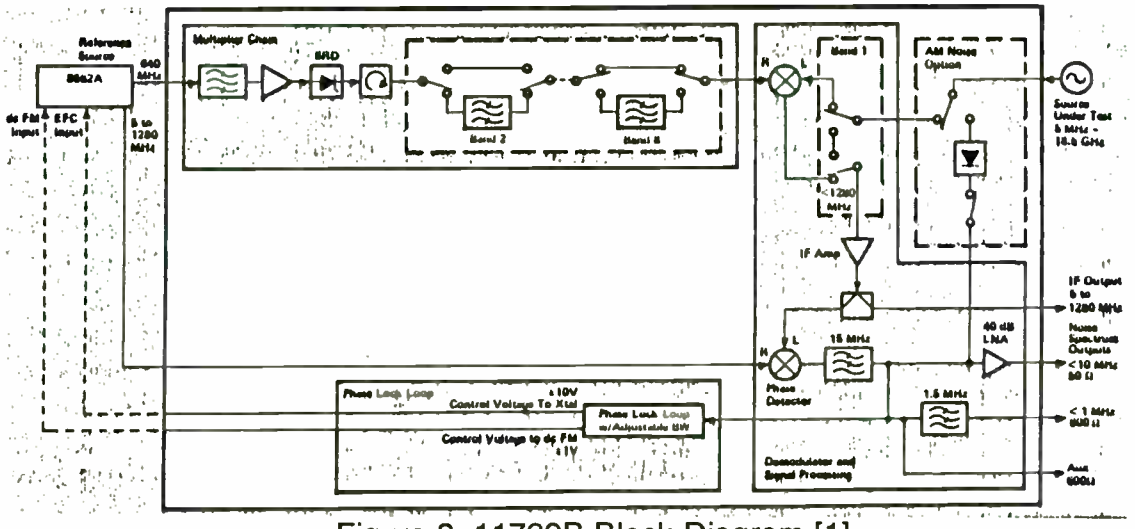


Figure 3. 11729B Block Diagram [1]

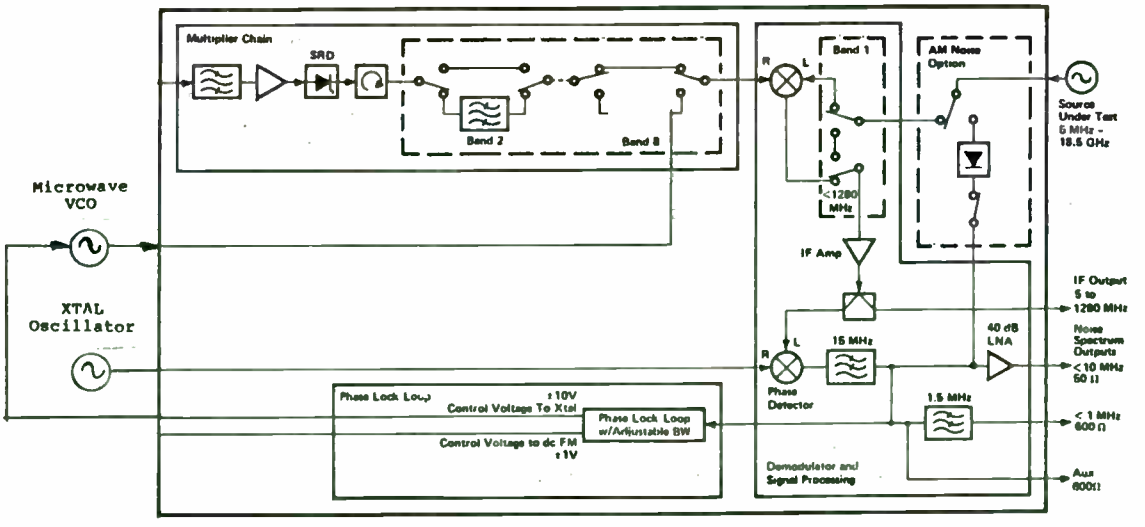


Figure 4. Modified 11729B Block Diagram [1]

The disadvantage to this technique is the fact that you must now have a low noise VCO as good as or better than the VCO inside the source under test. This also increases the uncertainty of the measurement accuracy since the contribution of the VCO isn't exactly known. It is a good idea to use a VCO that is either identical in design to the VCO inside the U.U.T. or one with known better phase noise.

The frequency of the crystal oscillator, and the offset frequency, is not critical but must be greater than 15 MHz in order to prevent saturation of the 40 db LNA inside the 11729. If you are so inclined, the 11729 can be further modified to add another filter of lower cutoff frequency in front of the LNA, thereby allowing a lower offset frequency.

Since the crystal oscillator is of much lower phase noise than the source under test, the phase noise it contributes is not very significant. Calibration and measurement of this configuration is now the same as that described in the product note "HP11729B-1" with the exception that you must add 3 dB to the noise for the effect of second VCO.

3a.) Frequency discriminator method with the HP11729.

The HP11729C with the built in 640 MHz oscillator can be utilized in the discriminator mode by following the procedure described in PN11729C-2 [2]. An alternative to purchasing the
current model of the 11729 is to purchase a used HP11729B and buy a good low noise 640 MHz source. A low cost approach to obtain the 640 MHz source is to purchase a low phase noise 160 MHz crystal oscillator and build the circuit in Figure 5 utilizing low cost doublers, amplifiers and a filter.

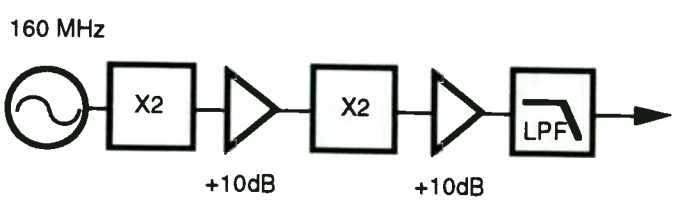


Figure 5. Low Noise 640 MHz Source

The discriminator method is accurate but can require a rather long delay line in order to obtain the required sensitivity. The HP product note PN11729C-2 [2] has a detailed discussion of the delay line discriminator.

Note that the 11729 makes it's measurement by looking at the noise outside the loop bandwidth. The loop is calibrated by unlocking and offsetting one source to produce a beat while attenuating the U.U.T. by 40 db to prevent saturation. Also the loop bandwidth correction factor is determined by injecting a source and recording the loop transfer characteristics. These are described in PN 11729B-1 [1].

Method 4. Using P.L.L. to Measure it's own phase noise. Compare the basic phase detector of diagram Fig. 6 to the diagram of a sampling P.L.L. Fig. 7.

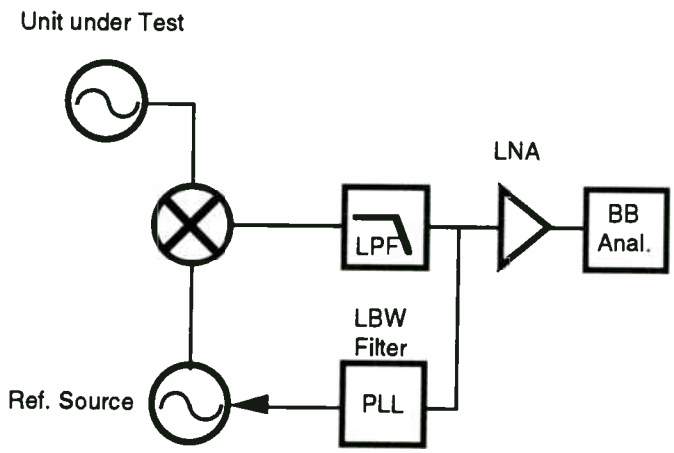


Figure 6. Phase Detector Technique

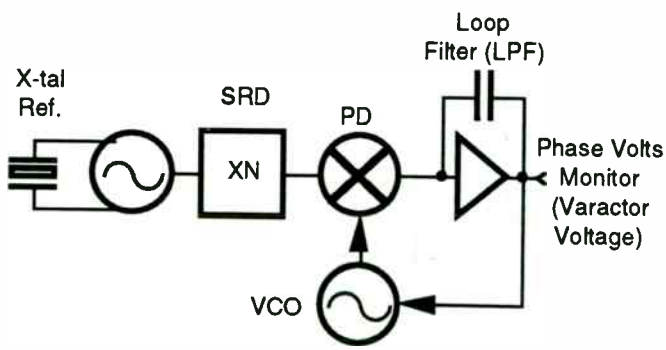


Figure 7. Sampling P.L.L. Diagram

Notice the distinct similarity in the functional block diagrams. This leads to the proposed technique of using the P.L.L. to measure it's own phase noise. Most P.L.L.'s have a test monitor point available after the loop filter to monitor the VCO voltage. This test point is very similar to the point used to measure phase noise in the test set except that it is inside the loop bandwidth instead of outside the loop bandwidth. This implies that it should be possible to monitor the noise at this test point with a baseband spectrum analyzer and determine the phase noise of the P.L.L. If the measurements are made within the loop bandwidth, the noise on this test point is the composite phase noise of the entire P.L.L.

This technique with no further correction is quite valuable for checking relative phase noise in a production environment. It is possible to simply clip lead a good F.F.T. analyzer to the PV test point of a P.L.L. and compare relative phase noise of several units. There are, of course, two basic assumptions; 1) the loop bandwidths are the same or sufficiently wide that the bandwidth doesn't affect the measurement, 2) the VCO constants are the same. The VCO constant can be determined by varying the reference oscillator enough to give $\pm$0.5V or less on the VCO curve, while observing the Δf on the output. This small delta measurement should be made about the nominal operating point.

$$Kv = \Delta f / \Delta V \ (MHz/V)$$

As long as the measurement frequency, f_m, is within the loop bandwidth, the phase noise can be calculated from the measured voltage and the VCO constant Kv. The absolute level of the noise voltage within the loop at the VCO input must be measured with a low frequency spectrum analyzer or wave analyzer. The Stanford Research SR550 is a F.F.T. analyzer that can measure the spectral noise density at an economical price. Wave analyzers such as the HP3581C series are also low cost methods of measuring the rms voltage at a specified frequency, f_m, away from the carrier.

Knowing the VCO constant, Kv, and having measured the rms voltage, Vrms, at an offset frequency, f_m, in a resolution bandwidth, BWres, the rms F.M. deviation can be calculated from:

$$\Delta f = Vrms * Kv * \sqrt{1Hz/BWres} \ [Hz] \quad (1)$$

Next the relative level of the FM carrier sideband ratio is calculated from the conventional FM theory. Obviously the voltage modulating the VCO is of a sufficiently low level to make the modulation index magnitude much less that 1, ($M<<1$). Therefore, the carrier to sidebands is defined by equation (2)

$$\frac{Psb}{Pc} = \frac{\text{Power density in the sideband}}{\text{Power of Carrier}} \quad (2)$$

(This is also the definition of phase noise, $\mathcal{L}(f)$.)

$$\mathcal{L}(f_m) = \frac{Psb}{Pc} = 20 \text{ Log } 10 \left[\frac{\Delta f}{\sqrt{2} * (f_m)} \right] \quad [3]$$

Where Δf = rms deviation in Hz
f_m = Noise modulation frequency in Hz.

With the measured voltage and modulation frequency, calculate Δf using equation (1) and using Δf and f_m, calculate $\mathcal{L}(f_m)$ using equation (2).

Example:

A 13 GHz PLO was tested by varying the reference frequency above and below the operating point. The VCO constant was measured as 0.935 MHz/volt.

The low frequency analyzer was connected to the VCO test point and the rms voltage measured at 10KHz as 393ηV/$\sqrt{Hz}$.

From (1)

$$\Delta f = 0.393 * 10^{-6} \text{ V} * 0.935 * 10^{6} \frac{Hz}{V} = 0.367 \text{ Hz}$$

From (2)

$$\mathcal{L}(10KHz)=20Log_{10}\left(\frac{0.367}{\sqrt{2*10^4}}\right)=-91.7\ dbc/Hz$$

Compare this to the HP 8566B analyzer reading of -61.6 dB in 1 KHz or -91.6 dBc / Hz (Reference Figure 9). Remember that this technique is valid only if the measurement is within the loop bandwidth. If it is not, the noise of the VCO outside the loop bandwidth will not be present as an error voltage within the loop.

Figure 8 is a direct plot from the SR770 F.F.T. analyzer in the Power Spectral Density (PSD) mode [4]. This mode selection automatically computes the resolution bandwidth with respect to 1 Hz. Note that this display is not a normal spectral amplitude display. It is a display of the rms frequency deviation versus modulation frequency.

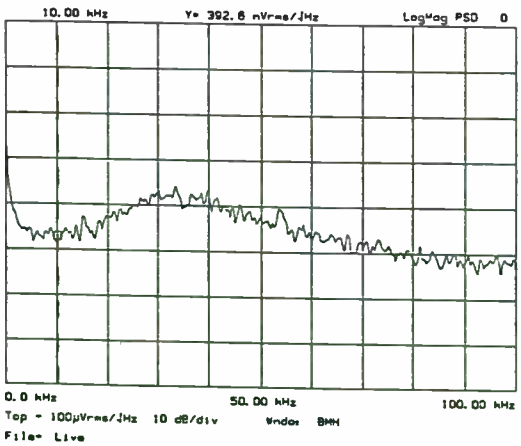


Figure 8. SR770 Spectral Density Display

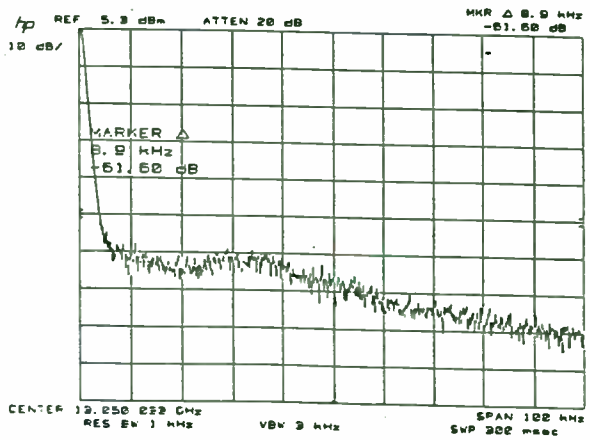


Figure 9. Hp 8566B Spectrum Display

This data doesn't become phase noise spectral information until equation 2 is applied to the data. One nice feature of the SR series analyzers is that it has a built in floppy disk on which the display information can be written in IBM ASCII format. This data can then be processed with a PC and appropriate software.

For comparative purposes, Figure 9 is a direct plot from a HP8566B spectrum analyzer with the carrier offset to the left and total span equal that of the SR 770. Several frequency points have been analyzed from Figure 9 and compared to Figure 10. This data indicates that the 10 KHz point is within the loop bandwidth of the P.L.L. and correlates well. The 50 KHz point is approximately 3 db off and the 100 KHz point 6 db off. This indicates that the measurements at these points are outside the loop bandwidth and must be ignored or compensated for.

Compensating for loop bandwidth effects require that the P.L.L. have an injection point that will allow measurement of the loop response to an outside stimulus.

If the P.L.L. manufacturer will cooperate with the purchaser, an additional loop injection test point can be provided to allow insertion of a test signal for measurement of the loop attenuation outside the loop bandwidth. Figure 10. shows a block diagram of the required loop with additional test point.

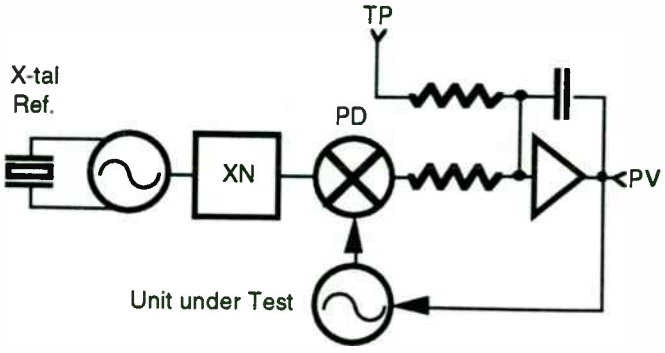


Figure 10. Modified PLO

Utilization of this test point will be in the same manner as described in PN 11729B-1 page 19 [1]. One obvious difference will be that the curve will have the opposite characteristic from Figure 4.14 in PN11729B-1 [1].

In conclusion, three conventional and one new method of phase noise analysis have been presented. The conventional methods require expensive equipment but yield accurate results. The newly proposed method takes advantage of the fact that the phase noise of a loop is present within the loop in a measurable form. The low cost aspect of the equipment required for measurement takes advantage of powerful processor technology applied in an affordable F.F.T. analyzer.

The author has utilized this technique for several years for relative comparison of production oscillators. Only recently has the technique been applied to an absolute phase noise measurement tool. It is hoped that others will conduct further research and analysis to improve the accuracy of the technique.

References

[1] Hewlet Packard Product Note 11729B-1 "Phase Noise Characteristics of Microwave Oscillators", Phase Detector Method, March 1984

[2] Hewlet Packard Product Note 11729C-2 "Phase Noise Characteristics of Microwave Oscillators", Frequency Discriminator Method, March 1984

[3] Sheilds, Robert B. , "Review of the Specification and Measurement of Short-Term Stability", The Microwave Journal, June 1969, pp49-55

[4] Stanford Research Systems "Operating Manual and Programming Reference for SR770 FFT Network Analyzer" ©1992

EXTERNALLY-INDUCED TRANSMITTER INTERMODULATION: MEASUREMENT AND CONTROL

Ernie Franke
E-Systems, ECI Division
P.O. Box 12248 MS 28
St. Petersburg, FL 33733

RF power amplifiers used for AM, FM, and TV broadcasting and for mobile, portable and base station radio transmitters are designed for maximum efficiency and output power, and not linearity. Intermodulation products can be generated in the final stages of power amplifier as a result of the mixing of the carrier frequency with one or more external signals coupled back through the transmitting antenna. Industry uses several terms such as reverse intermod, back intermod, and antenna-induced intermod to describe this externally-induced transmitter intermodulation. The level of these intermod products may be characterized in the laboratory, by injecting a known swept-frequency signal level into the output port of the power amplifier, and then used for radio system interference analysis. Although straight forward in concept, the measurement procedure and the interpretation of the results of intermod testing often cause considerable misunderstanding.

With the advent of wideband transistor power amplifiers and the need to operate at crowded repeater sites, especially with frequency-agile synthesizers, the characterization of broadband, externally-induced intermod performance becomes more important. Transistor amplifiers provide a wideband match allowing spread spectrum, fast tuning, and frequency hopping operation. This is required for mobiles operating in military tactical or cellular mobile telephone environments. Vacuum tubes have intermod performances similar to transistors, but tube power amplifiers typically operate in narrowband, cavity or high-Q output circuits which offer rejection both to the incoming interfering signal and to the generated intermod product.

This article describes how to examine externally-induced intermod performance using a wideband technique for characterization of modern power amplifiers. Variables such as supply voltage, transistor operating point, etc. are examined for effects on intermod performance. Methods of improving intermod performance through circuit design and through the use of external devices are also investigated. With this insight into the intermod process, the confusing measurements in the field or in the lab may be better understood and the intermod performance improved.

INTERMOD INTERFERENCE

Intermodulation (intermod) products can be generated in the final amplifier stages of radio transmitters that are in close proximity with other transmitters, receiver front-ends or mixers in proximity with several strong signals, common antenna feed systems, and rusty or corroded metallic structures that are exposed to high RF fields. A nonlinear amplitude response in the transmitter output power amplifier, an over-driven receiver front-end, an oxidized antenna or a corroded cable in a multiple-transmitter location can all give rise to intermod interference which masks desired weaker signals at the receiver. Although receiver and metallic intermod may be present in such an environment, the major source of interference is due to externally-induced transmitter intermod, and will thus be the thrust of this article.

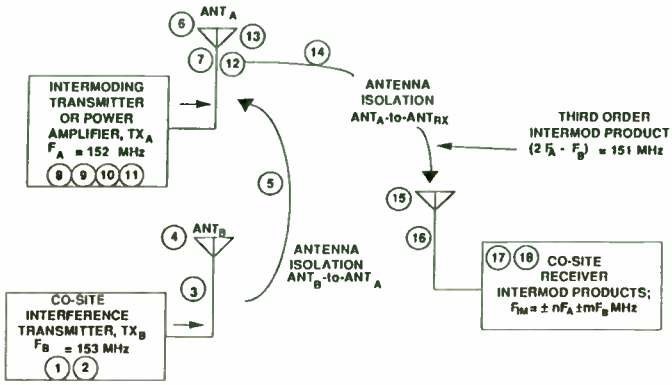


FIGURE 1 Co-site transmitter intermod interference is caused by mixing action in the output stage of an RF power Amplifier. Circled numbers represent gains and losses from the intermod budget below.

RF Budget for computing externally-induced transmitter intermod interface

TABLE 1

Location	Equipment	Frequency	Description	Level
1	Co-Site TX_B	F_B	Co-Site Interfering Power	+50 dBm (100 Watts)
2	Co-Site TX_B	F_B	Output Filter Cavity Loss	-1 dB
3	Co-Site TX_B	F_B	Transmission Line Loss	-2 dB
4	ANT_B	F_B	Antenna Gain Toward TX_A	+3 dB
5	TX_B-to-TX_A	F_B	Space Loss from ANT_B to ANT_A	-2 dB
6	ANT_A	F_B	Antenna Gain in Direction of TX_B	+5 dB
7	TX_A	F_B	Transmission Line Loss	-15 dB
8	TX_A	F_B	Output Filter Cavity Loss	-1 dB
9	TX_A	F_B	Externally-Induced Interference Power into TX_A Power Amplifier	+20 dBm (0.1W
10	TX_A	F_B-to-$(2F_A - F_B)$	Third-Order Intermod Performance of Power Amplifier	-20 dB
11	TX_A	$2F_A - F_B$	Output Filter Loss $(2F_A-F_B) = F_B$	-1 dB
12	TX_A	$2F_A - F_B$	Transmission Line Loss $(2F_A-F_B) = F_B$	-2 dB
13	ANT_A	$2F_A - F_B$	Antenna Gain in Direction of Co-Site RX	+6 dB
14	ANT_A -to-ANT_{RX}	$2F_A - F_B$	Space Loss	-90 dB
15	ANT_{RX}	$2F_A - F_B$	Antenna Gain Toward TX_A	+7 dB
16	Co-Site RX	$2F_A - F_B$	Transmission Line Loss	-2 dB
17	Co-Site RX	$2F_A - F_B$	Input Filter Cavity Loss	-1 dB
18	Co-Site RX	$2F_A - F_B$	Interfering Signal Input at RX	-83 dBm (16μV

Intermod interference generated in a transmitter output stage typically occurs at sites where transmitter antennas are located in close proximity to each other and mutual coupling exists between these antennas, Figure 1. The power radiated from a co-located interference transmitter (TX_B) may be coupled into the antenna of the transmitter in use (TX_A). It must be emphasized that both the transmitter under inspection (TX_A) and the interference station (TX_B) are both operating properly and if either were operated alone they would not generate intermod products. Many installations which share the same mast often bundle transmission lines together in a trough, where the signal from one transmitter couples through the

braid of the coaxial cable into the output port of another transmitter. This results in interfering signals entering the final amplifier stage of the intermoding transmitter or power amplifier. Since most transmitters operate in either the Class C, D or E mode for maximum efficiency, the power amplifiers are non-linear and hence act as mixers. The desired carrier signal frequency (F_A) mixes with the interfering signal (F_B) to produce intermod products (F_{IM}), which are then re-radiated via the transmitter antenna (ANT_A), just as is the desired signal, and are received as interference signals by nearby receivers.

INTERMODULATION PRODUCT FREQUENCIES

Let's begin by defining externally-induced transmitter intermodulation distortion. It is simply the mixing of frequencies in the final transmitter stage to generate new frequencies, which are then radiated from the transmitting antenna. This is different from the intermod distortion performance of a linear power amplifier used for single sideband or digital data service where two-tone test signals are injected at the input of the power amplifier stage. If two signals are present in a non-linear device, such as in the output stage of a power amplifier, mixing will occur, creating additional spurious signals according to the simple sum and difference mixing formula:

$$F_{IM} = \pm nF_A \pm mF_B$$

where; F_{IM} = frequency of the intermod product

F_A = frequency of the intermodulating transmitter

F_B = frequency of the co-site interference signal

n, m = 1,2,3 . . . . integers

If more than two frequencies are involved, the number of combinations rises rapidly. The order of the intermod product is equal to the sum of the integers n plus m. The most important intermod products are those that are closest to the carrier frequency with low integers, because they are both the strongest and the most difficult to filter. The third-order products ($2F_A - F_B$ and $2F_B - F_A$) and the fifth-order products ($3F_A - 2F_B$ and $3F_B - 2F_A$) are shown in Figure 2. The amplitude of each product is shown relative to the output signal carrier level (dBc). The third-order intermod performance is simply the difference between the co-site interfering signal level and the largest 3rd order product. The intermod products are spaced at the difference frequency ($F_B - F_A$).

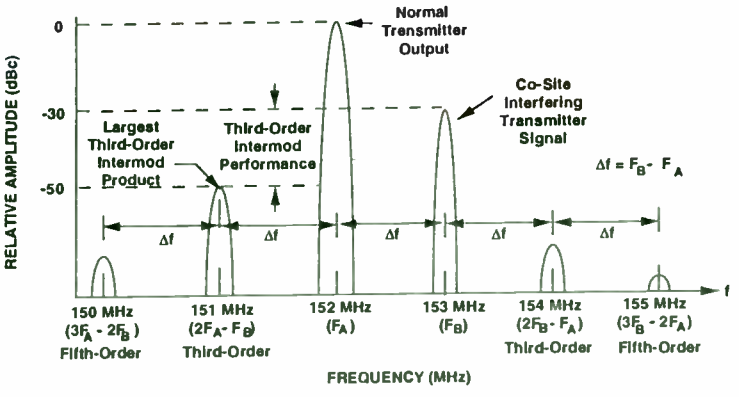


FIGURE 2 A nearby interfering transmitter signal F_B mixes in the non-linear output stage of a power amplifier with the second harmonic $2F_A$ of the operating frequency to form a strong third-order intermodulation product at $2F_A - F_B$.

In a typical transmitter output circuit, as shown in Figure 3, a co-site interfering signal (F_B) is coupled into the antenna from a nearby transmitter and beats with the second harmonic of the operating frequency ($2F_A$) in the collector or drain circuit. The third-order, sum intermod product ($2F_A + F_B$) is blocked from the antenna by the lowpass filter (LPF), normally present in a transmitter to reduce harmonic energy. The third-order, difference intermod product ($2F_A - F_B$) however will readily pass through the LPF and be radiated by the antenna. If the interference frequency is close to the operating frequency, even a high-Q cavity filter will not reflect the interfering frequency or filter the intermod product.

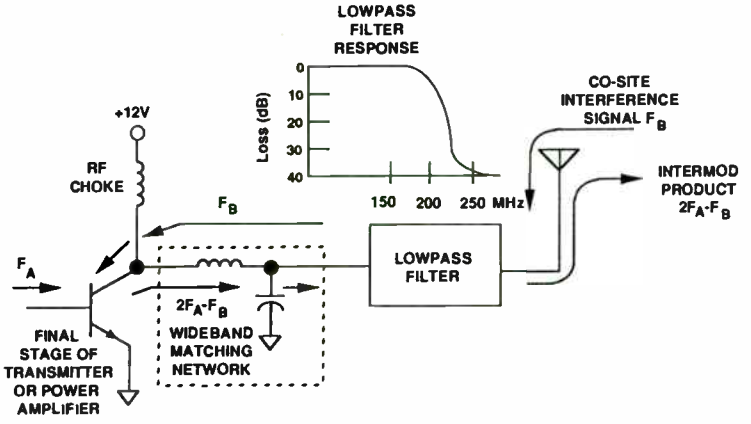


FIGURE 3 The third-order intermod product ($2F_A - F_B$) easily passes through the output lowpass filter if F_A and F_B are reasonably close to the same frequency.

INTERMODULATION PRODUCT POWER LEVEL

Intermodulation products are not only frequency related, but amplitude related as well. Because the second harmonic power varies 2 dB for each 1 dB change in output power, the intermodulation signal power for the ($2F_A - F_B$) third-order intermod product also varies 2 dB for each 1 dB of variation in normal transmitter output power. The intermod product power however varies directly dB for dB with the incoming interference signal power for this product. The signal level of the weaker ($2F_B - F_A$) intermod product however varies 2 dB for each 1 dB of variation in interference power. This power relationship is helpful for discerning the exact intermod culprit. This power relationship is shown in Figure 2 for close-in products, where the interfering signal is injected 30 dB below the normal transmitter output and the resultant intermod product is 20 dB below this interfering signal level. Class C power amplifiers typically display a third-order, externally-induced intermod performance between 15 and 30 dB. Higher-order intermod products will drop 15 dB for each successive odd higher order. The 5th order products are typically 15 dB lower than the third-order products and the seventh-order products are approximately 30 dB down from the third.

CO-SITE INTERMOD BUDGET

A typical co-site environment involving a 100 watt transmitter (TX_B) coupled into a transmitter (TX_A) creates a third-order intermod product ($2F_A - F_B$), as itemized in Table I. Because the interfering transmitter frequency (F_B) is very close to the desired operating transmitter frequency, the transmission line

losses and filter losses at each site will be approximately the same for the desired operating, interfering and intermod frequencies. The antenna gain at each frequency will however not necessarily be the same because of directionality. It is interesting to note that the desired output power level of the transmitter under question (TX_A), which generates the intermod products, does not enter into the equation. The transmitter power level is always assumed to be much larger than the co-site coupled interference power level, and simply acts as a pump for generating mixer intermod products.

The resultant intermod product signal level of -83 dBm at the input to a nearby sensitive receiver will most likely overpower the desired signal. To minimize receiver interference, the amplitude of the intermod product needs to be on the order of the noise figure of the receiver. Thus even -20 or -30 dB of intermod performance at the output stage of a transmitter itself is inadequate for receiver interference protection. Control of intermod interference takes the combined effort of power amplifier design, band-pass and band-reject filtering, antenna separation, circulators, and frequency planning to effectively control intermod interference.

TRACKING DOWN ON-SITE INTERMOD INTERFERENCE

The first step in any intermodulation analysis is to write out a table of all frequencies present at a site and then compute the intermod products. Next it is essential to determine whether the intermod product originates at the transmitter, receiver or a combination of both. A variable RF attenuator placed at the receiver input permits observation of any overloading effects. An intermod product, which is produced due to receiver overload, will drop drastically as soon as the signal causing the overload is attenuated below the overload point. If the intermod product drops at the same rate as the introduced attenuation, then the receiver is not the cause and each co-site transmitter must be investigated. If the intermod product drops by twice the attenuator setting, the problem is in the receiver and the troublesome product is second-order. Third-order products in the receiver drop 3 dB for each 1 dB change in front-end attenuation.

Transmitter intermod is most easily identified by turning off suspected transmitters, changing antennas, changing operating frequencies, or operating transmitters into dummy loads. If only the interfering power is decreased by 10 dB, the ($2F_A$ - F_B) internal product will reduce by 10 dB, while the other third-order ($2F_B$ - F_A) product will drop by 20 dB. This reduction in signal level should be reasonably close (±3 dB) to this predicted value. If two FM voices are heard simultaneously at normal levels, it is probably second-order intermod. If the voice on the intermoding transmitter channel sounds loud and distorted, the intermod is likely to be third-order ($2F_A$ - F_B), or possibly fifth-order intermod. The deviation of the carrier signal F_A is multiplied, whereas the deviation of interference signal F_B remains constant.

MEASUREMENT SET-UP

Externally-induced intermod performance is best described, not by a third-order intercept point as is done for small-signal linear amplifiers, but by a third-order intermodulation distortion (IMD) measurement using a carrier injected into the power amplifier output port. The performance is always stated for the third-order ($2F_A$ - F_B) product, because it is the largest product. The device under test can be a single transistor in a fixture containing input/output matching, an amplifier module, or an entire high-power transmitter.

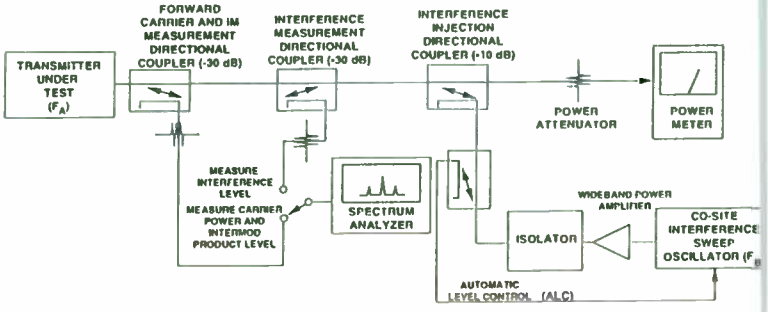


FIGURE 4 A continuous curve of intermod performance, including close-in measurements, may be made using the wideband test set-up.

A wideband equipment arrangement [1] for measuring the externally-induced intermod performance of transmitter (TX_A), Figure 4, shows a directional coupler for measuring both the forward carrier power at the operating frequency (F_A) and the intermod product levels. Two more directional couplers are added to inject the level of the co-site transmitter power at the interference frequency (F_B) and to display the actual injected value on the spectrum analyzer. The broadband co-site interference signal is injected using a broadband isolator to avoid intermoding in the sweep power amplifier.

The wideband intermod test set-up is used to plot the intermod performance as a continuous function of frequency separation on both sides of the carrier frequency. The spectrum analyzer is configured for a maximum hold display, while the interfering frequency is only swept on one side of the carrier. Because the third-order ($2F_A$ - F_B) internal product, which shows up on the opposite side of the carrier from the injected signal, is the strongest intermod product, it is always the resultant intermod performance curve recorded by the peak hold display function. A plot is performed before the interference signal is swept on the other side of the carrier and then plotted again. By selectively lifting the plotter pen, a plot of only intermod performance is obtained. The resultant display of a typical 150 watt, push-pull UHF power amplifier yields a clear picture of third-order intermod performance , Figure 5, especially for close-in products. The horizontal axis shows the injection frequency relative to the output carrier while the vertical axis shows worse-case third-order intermod energy. The externally-injected interfering frequency is always on the opposite side of the carrier from the intermod energy.

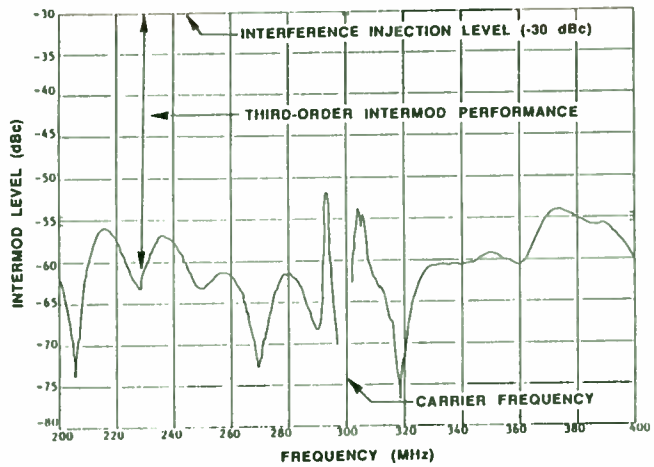


Figure 5 By sweeping the injected interference frequency, a clear picture of close-in, third-order intermod performance is produced. In this case it varies between 25 and 40 dB.

If the interference is so close to the carrier that a bandpass filter cannot provide the required selectivity, a ferrite isolator must be used. An isolator is a circulator with the third port terminated. The circulator, Figure 6, is a device which uses the interesting gyromagnetic behavior of ferrite. A transmission line is sandwiched between two ferrite discs inside a static magnetic bias. The 3-port circulator has the property that RF incident at port 1 emerges at port 2, a wave incident on port 2 emerges at port 3 and a wave incident at port 3 emerges at port 1 in a cyclic or circulating manner. An interference signal coupled into the antenna will be shunted to the circulator termination. Approximately 20 to 30 dB of reverse isolation is achieved for each circulator section added.

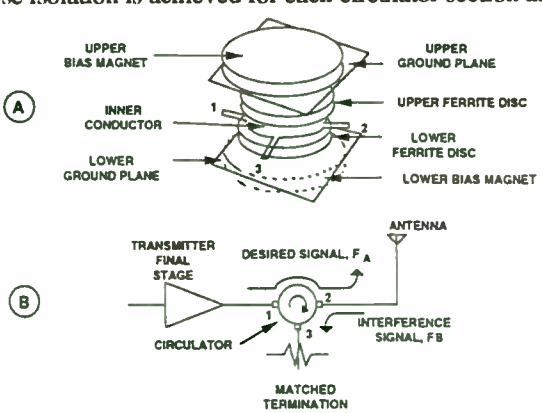


FIGURE 6 Interaction between the magnetic field of the RF signal and the magnetically biased ferrite produces a tendency for the signal to follow a circulating path.

CLOSER EXAMINATION OF INTERMOD GENERATION

The intermod products present at the output circuit of a power amplifier are dependent on several items; the level of injected interference signal, power amplifier output return loss, amplifier linearity, operating frequency, spectral proximity of the interfering signal, reverse isolation of the final stage, bias circuitry, supply voltage, and power combining techniques.

Third-order intermod performance improves for example, Figure 7, as the RF output power level is decreased or as the power supply voltage is increased. The intermod performance consistently improved by 10 dB for each 5 Vdc increase in power supply voltage. The varactor effect of an RF voltage modulating the output capacitance becomes a smaller part compared to the steady state component. It also improves as the required output power is backed off from the design value, which is 70 watts for the example shown. At lower power, the third-order intermod power varies 2 dB for each 1 dB change in output power level. At higher power levels the intermod performance degrades 3 to 5 dB for each 1 dB change in power.Typically the final stage in an FM transmitter is operated in class C very close to its saturated power level for maximum efficiency. Distortion due to saturation may be decreased if the size of the output transistor is increased or if the output power is turned down, usually with a slight sacrifice in efficiency.

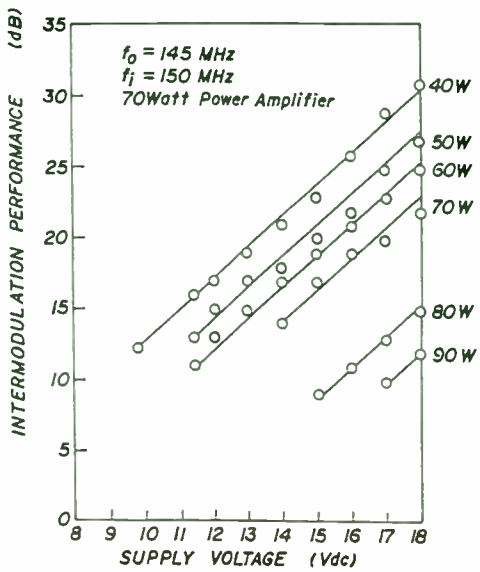


Figure 7 Third order intermod performance improves as the power level is decreased or as the power supply voltage is increased.

Measured externally-induced intermod performance may also be partially composed of forward intermod products. The collector-to-base or drain-to-gate capacitance is typically large enough and the base or gate impedance levels are high enough that reverse transmission values of -10 to -20 dB are common. Thus the interfering signal appears at the input to the final stage and mixes in the same manner as in forward intermodulation measurements for linear power amplifiers. Signals reflected off interstage filters, several stages before the final output stage, have been observed to affect third-order intermod performance, as shown in Figure 8.

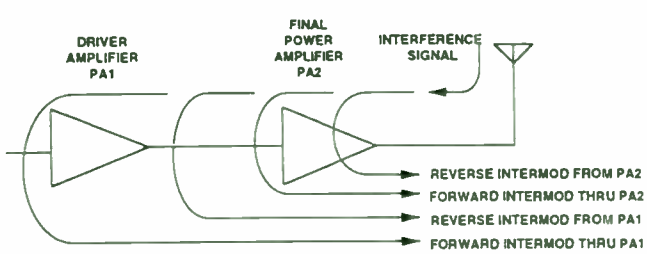


FIGURE 8 The intermod perfomance of a power amplifier is a composite of the forward and reverse intermod of the driver and final amplifiers.

The intermod performance or turn-around loss of a power amplifier is related to the level of the interfering signal, frequency spacing between the transmitter carrier frequency and interfering frequency, the bandwidth of the transmitter, and the value of loading. It is not only a function of the power amplifier output circuit bandwidth, but also of the input circuit. An experimental arrangement for measuring reverse isolation is shown in Figure 9.

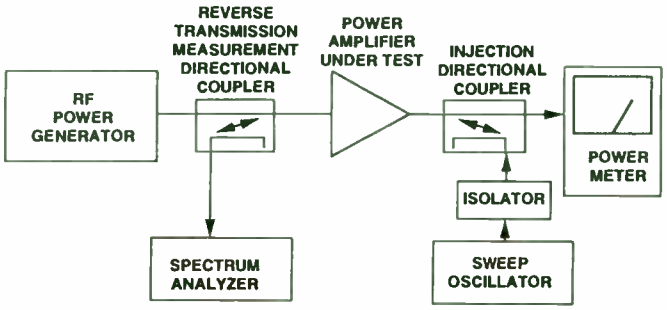


FIGURE 9 The reverse isolation of power amplifier may be measured under actual operating conditions by injecting a sweep signal into the output port and monitoring the input.

POWER AMPLIFIER INTERMOD IMPROVEMENT

Bias network impedance can also influence intermod performance. Beat notes between the operating frequency and the interfering frequency can be formed both at the output and at the input of a power amplifier. These low frequency difference frequencies can then modulate the operating frequency and appear on the spectrum analyzer in the same manner as third-order intermod products. Solid-state amplifiers typically have a much higher gain at low frequencies, as seen in Figure 10 for a typical VHF bipolar transistor. The power gain expression contains a $1/f^2$ term indicating that the power gain decreases at 6 dB per octave with increasing frequency. This figure displays the upper bound of available gain for a common emitter transistor whose input and output are matched. The unity gain intercept frequency, f_T, is a projected point due to the effect of package parasitics coming into effect as the intercept is approached. This slope approaches the low-frequency beta (H_{fe}) at f_B.

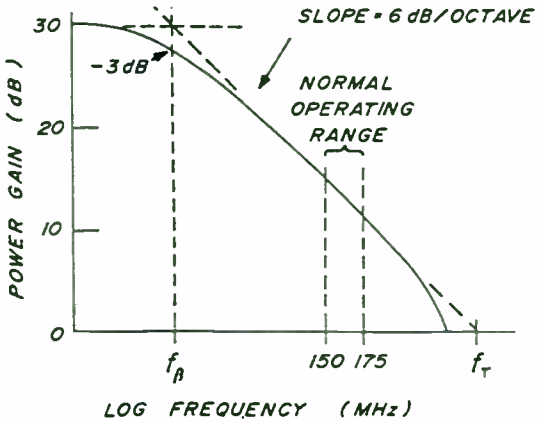


FIGURE 10 Solid state amplifier have excess low-frequency gain and thus enchance close-in mixing product.

The techniques of using feedback and shunting networks [2] may be used to decrease this low-frequency gain and improve power amplifier stability. The first method of employing a network to provide feedback is shown in the negative feedback circuit of Figure 11A. The capacitor is used to block the collector supply voltage from reaching the base. The series inductor and the resistor are used to provide 6 dB/octave of negative feedback to offset the effect of the in transistor gain slope.

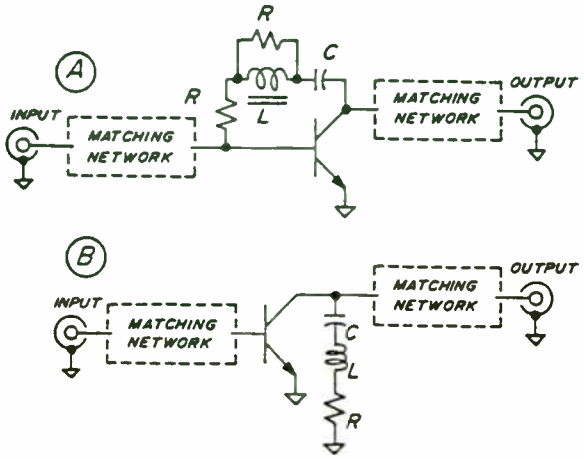


FIGURE 11 The large low-frequency gain may be controlled through A.) negative feedback or B.) shunt loading.

Low frequency gain may be controlled and stability improved by negative feedback but the intermod performance is typically degraded, as shown in the single-sided intermod sweep of Figure 12 for a 150 watt, UHF power amplifier. The second method of low-frequency gain reduction involves collector loading, as illustrated in Figure 11B, where an RLC circuit has been attached to the collector. The circuit is transparent at the operating frequency due to the inductor. At low frequencies the inductive reactance decreases such that the loading resistor appears directly across the collector. The capacitor is added solely to block the dc supply voltage.

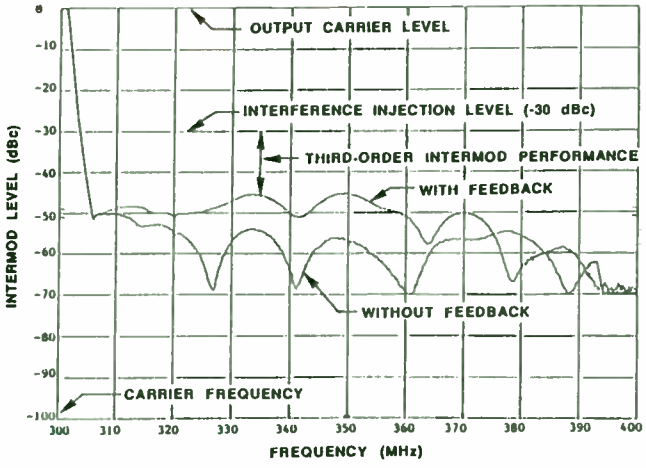


FIGURE 12 The third-order intermod performance is degraded when low-frequency feedback is added to improve stability.

An alternate method of increasing stability is achieved by low-frequency, shunt-loading of the base and collector as shown in Figure 13. The base bias current return circuit consists of two chokes, L1, which is a small inductor in the range of 100 nanohenries chosen to function as a choke at the lowest operating frequency and L2, which presents a high impedance down to very low frequencies. At normal operating frequencies, the base is effectively isolated by inductor L1 from resistor R1, forming a low Q arrangement. The Q of the network is given as $Q = 2\pi f L1/R1$ which is low. The resistor R1 is chosen to be 10 to 20 ohms. At lower frequencies, where the power gain is much greater, the shunting effect of R1 increases due to the lower inductive reactance of L1. The larger value inductor L2, consisting of a few turns of wire through a powered-iron bead, is a low ohmic dc path across R1, Providing a dc bias for the base and a high impedance at low frequencies. Thus the base is effectively resistively loaded by R1 at low frequencies, where the power gain is greatest.

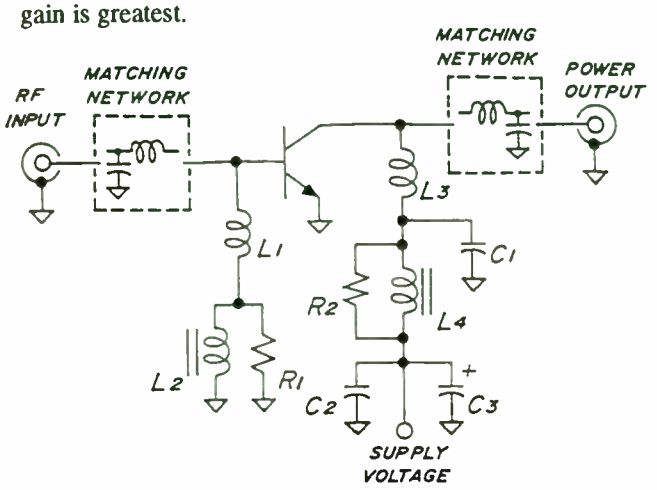


FIGURE 13 Low frequency loading techniques applied to the base and collector feed networks.

The collector feed network accomplishes a similar feat by resistively loading the collector at frequencies below the operating band. Within the frequency range of normal operation the small inductor L3 and the small shunt capacitor C1 act as an L-network to transform the resistance of R2 to a high value so that it will not load the collector. At the operating frequency, the powered-iron loaded inductor L4 appears as an open, while the capacitor C2 acts as a very low impedance. Below the normal operating range the input impedance to the network looks resistive asymptotically approaching the value of R2, typically 10 to 20 ohms. The wattage of this loading resistance need only be 1/2 or 1 watt because it is designed to absorb less than 1/4 watt of signal power at the operating frequency. The bypass capacitors, C2 and C3, placed at the supply voltage consist of a disc ceramic for bypassing at RF and an electrolytic for bypassing at audio frequencies. The resultant intermod performance improvement with base and collector low-frequency, resistive-loading is shown in Figure 14.

As the output power from a VHF or UHF solid-state transmitter is increased beyond approximately 50 to 75 watts, where the power handling capacity of a single device is exceeded, the output stage must combine the power from two transistor packages to achieve power levels up to the 100 to 150 watt level. If the output stage uses a push-pull arrangement, it will display better third-order intermod performance because second harmonic energy is decreased.

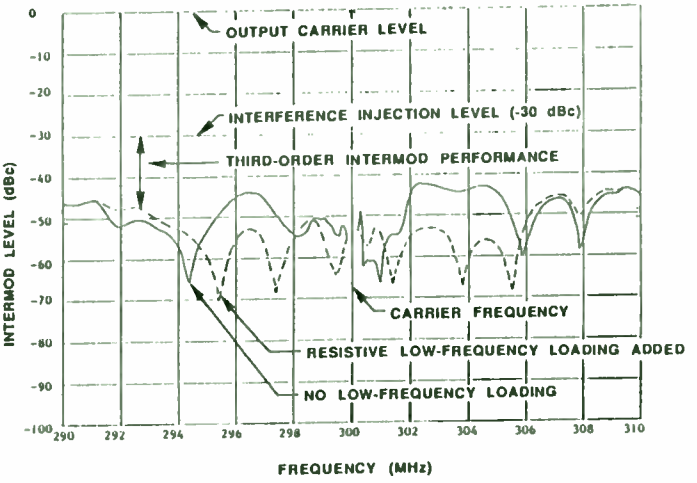


FIGURE 14 The third-order intermod performance is improved with resitive low-frequency loading techniques added to the base and collector feed networks.

In push-pull operation the even harmonic (second, fourth, etc.) distortion is balanced out. There are also several circuit techniques that may be applied to the output stage of the amplifier which will improve performance of a transmitter. These techniques include operating the final push-pull, using 90° hybrid combiners [3], and using larger output devices.

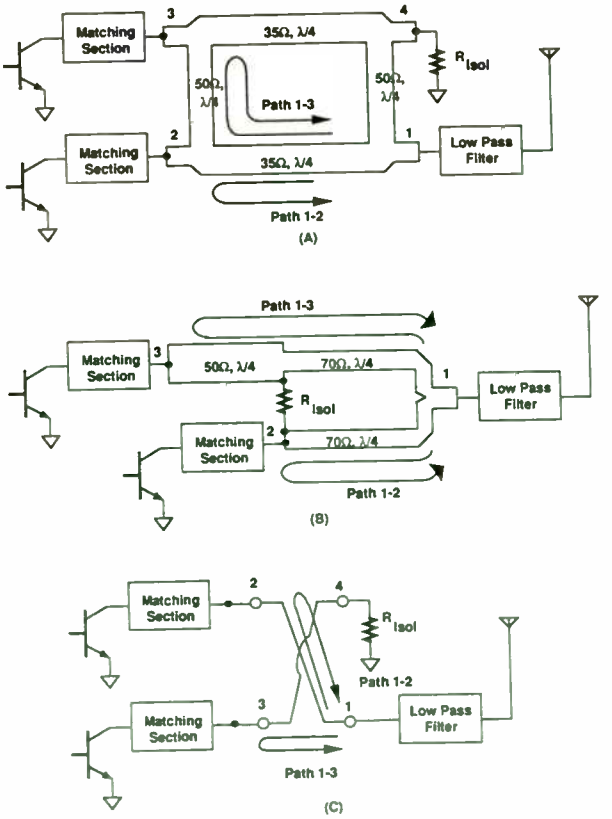


FIGURE 15 (A) Branch line coupler (B) staggered Wilkinson combiner and the (C) coupled - stripline coupler reduce externally-induced intermod generation.

The use of several 90° hybrid combiners is shown in Figure 15. The cancellation is strictly a property of the circuit configuration and is independent of the transistor linearity. For each type of combiner the round-trip in path 1-3 is half-wave greater than that of path 1-2. Thus the intermod product created at each collector arrives at the summing (port 1) 180° out-of-phase with each other and thus will cancel out. Maximum cancellation will occur at the center frequency and will fall off as the transmission lines are no longer quarter-wavelength at the operating frequency. The coupled-stripline quadrature hybrid however maintains a 90° relationship for nearly an octave bandwidth. Under practical conditions, where the two amplifiers are fairly well matched, a 10 to 30 dB improvement in third-order intermod performance can be expected.

REFERENCES

1.) E. A. Franke, "Test Setup Gauges Externally - Induced Transmitter IM", Microwaves and RF, April 1993, pp. 95 - 98.

2.) E. A. Franke and A. E. Noorani, "Improving the Stability of the Mobile Radio Power Amplifier", Proc. of the 33rd IEEE Vehicular Technology Conference, Toronto Canada, May 24-27, 1983, pp. 228-234.

3.) E.A. Franke and A. E. Noorani, "Wilkinson Hybrid with Quarter-Wave Offset", RF Design, November/December 1981, pp. 32 - 38.

A 3 GHz 50 Ohm Probe for PCB Measurements

Joel Dunsmore
Hewlett-Packard

Robert Kornowski
Motorola

Chuck Tygard
Everett-Charles Tech.

Introduction:

With the trend to commercial use of RF and microwave circuits, and the large increase in the communication segment, the use of PC boards as RF circuits demands non-traditional techniques to provide accurate yet automated measurements of these circuits. Manufacturing environments require high through-put measurements, and improving quality standards demand finer limits on measured parameters. An RF probe designed by Motorola and cooperatively developed with Everett Charles Technologies can be combined with high performance Vector Network Analyzers (VNAs) to support these requirements in automated measurements such as gain, return loss, output power and output match.

RF systems often use PC boards as interconnects between more complex modules, such as frequency converters, filters or amplifiers. These components may need to be pre-tested before assembly, and built-up subassemblies need to be measured before final assembly. Also, the RF match of a sub_assembly may be measured at the "socket" to ensure it is good before an expensive module is added. For example, the match of a tuning section may be measured before a power amplifier transistor is soldered in place. The problem of making the connection for the measurement, especially in an automated board test environment, requires improvements in probes and error correction in the measurement instrument.

PC Board RF Measurements:

Requirements for measurements of RF characteristics on PC board assemblies vary with the type of measurement. The predominant error to overcome is the match or return loss of the probe-to-board interface. For gain and power measurements, this mismatch may have only small effects, but for return loss or impedance measurements, the mismatch of the probe-to-board interface may be greater than the return loss of the component on the board. This mismatch adds directly to the directivity, source match, and frequency response of the test system.

Even if the test system performance is better than the match being measured significant errors in the measurement may occur. For example, figure 1 (a) shows a test circuit consisting of a 50 ohm resistor in series with a 50 ohm line, measured with a system having an interface to the test circuit of 19 dB return loss. Ideally, the test circuit has 100 ohms input impedance (about 9.5 dB return loss), and a voltage loss of 0.5 (about -4 dB insertion loss). Figure 1 (b) shows the measurement of the system interface having 19 dB return loss at 3 GHz. Figure 1 (c) shows the return loss of the test circuit without the degradation of the test system interface

The combination measurement of gain and return loss are shown, with and without the error of the measurement probe in Figures 1 (c) and 1 (d). The gain is in error by about .2 dB (theoretical worst case is .3 dB), but the match is in error by about 3.7 dB (theoretical worst case is 3.8 dB). Clearly, measurements of return loss depend very heavily on the match of the measurement system.

There are several factors which make RF PC Board contacts difficult. Most test equipment will be connected using coaxial cables, terminated in common connectors such as SMA, SMB, type N, or BNC. It is difficult to to make a good, low impedance ground connection to the PC board. Also, it is difficult to maintain a constant impedance on the RF signal line with respect to ground. Adding solder-on connectors to PC boards is expensive, and does not solve the automated test problem. Additionally, most non-solder probing techniques suffer from errors introduced by probe placement uncertainties. Even a good probe, probing a well controlled line structure will make a poor measurement if it is not exactly placed.

A New Kind of Probe - The K-50:

A new probe developed for automated testing at Motorola is well suited for RF PC board testing, shown in figure 2. It consists of an SMA connector which maintains constant impedance until terminated in a spring-loaded contact assembly. The grounds for the probe are brought down from very near the center contact with two additional spring-loaded contacts. The crown of the contact is designed to provide a small air space, more nearly maintaining a constant 50 ohms impedance to a footprint on the PC board. It is clear that this probe solves several of the problems mentioned above.

The grounding scheme of this probe has been found to provide a low impedance and be resonance free up to 3 GHz. The spring probes ensure positive contact even if not perfectly aligned either vertically or laterally. The symmetric, transmission line nature of the probe tip design provides for constant impedance, with good match beyond 3 GHz. Finally the rugged design combined with removable spring contacts provides ease of maintenance and reliability necessary for high volume automated testing. This probe demonstrates the first non-solder RF connector compatible with automated board testing equipment.

Enhancing the Probe with Vector Network Analyzers:

The vector network analyzer (VNA) can be used to make many of the necessary measurements for testing RF PC boards. It is ideally suited for gain, power and match measurements, as well as power gain compression, harmonic distortion, impedance vs. frequency and impedance vs. distance. A key to success when using VNAs is in properly correcting for errors due to mismatch in the instrument or connections to the circuit. The process of removing systematic errors is sometimes referred to as calibration, and involves the measurement of precisely known standards to determine these systematic errors.

With normal calibration techniques, the errors must be stable, and cannot be corrected if they change from measurement to measurement. When using the probe, the connection repeatability error due to probe placement changes cannot be corrected in the normal fashion. However with a technique known as time domain gating, the effect of these non-repeatable errors can be removed from the measurement. This is especially useful in measurements of match or return loss where the PC board device has a match as good as or better than the uncorrected performance of the probe.

Probe Details:

The mechanical construction of the probe affords both good RF performance and rugged, reliable testing life. The key to this is combining the precision spring contacts used in high volume PC board testing with the good RF performance of the SMA connector.

Figure 3 shows some of the mechanical details of the probe, including the replaceable spring loaded contacts.

The electrical performance of the probe can be described by its measurement of a low return loss load, Figure 4. This measurement was made by calibrating at the SMA connection, and measuring the probe and PC board load. When properly matched to a PC board trace, the probe can maintain a 26 dB return loss to 1 GHz, and a 20 dB return loss to 3 GHz. For calibration purposes, the probe has an open circuit capacitance of 0.4 pF, and a compressed short circuit length of about 6 mm. Here, compression means placing the probe on a piece of bare PC board material, and fully compressing the spring contacts. The short circuit length of the probe results in 36 picoseconds delay. Figure 4 (a) shows a Smith Chart, from which we can obtain the parasitic capacitance of the probe. Figure 4 (b)

shows the magnitude of the return loss in dB, with markers placed at 1 and 3 GHz.

To obtain best results with the probe, the PC trace must be properly configured to accept the probe. This should pose no problem in new designs, as the probe configuration is common to many devices now supplied to the RF market. The probe pad is shown in Figure 2; it is important that the PC trace "tail" is controlled, as it appears as a open circuit stub on the line.

Use with a VNA:

To use the probe with a VNA, it is first necessary to perform some type of error correction. For gain or power, it is only necessary to provide some source of RF at (a known level) to an SMA to PC board adapter. Placing the probe on the PC board, and performing a trace normalization (sometimes called a "thru cal") will correct for any gain slope or frequency response in the probe, connecting cables, or network analyzer. If the mismatch of the connecting cables and network analyzer is large, a small value attenuator can be added between the probe and the cables.

Error correction is much more important in measurements of return loss, input match, or impedance. The simplest form of correction is a standard open/short/load calibration in SMA (known as 3.5 mm for precision standard), adding the probe after calibration. Here, any mismatch in the probe interface is ignored. This will correct for any mismatch in the network analyzer and cabling to the probe. Since the probe has very good RF performance, this may be sufficient for many measurement needs.

In some cases, such as using the probe to characterize the impedance of PC board components, e.g. modeling active or passive components, a better calibration is needed to provide the correct phase reference. This can be obtained by starting with the normal calibration, then resetting the phase reference by shorting the probe, and adding electrical delay (also called electrical port extension) until a 180 degree phase trace is obtained.

Alternatively, a calibration "kit" can be made consisting of a PC board open (bare PC board), a PC board short, and a PC board load. Placing the probe on the three standards, as called for in the calibration routine, will correct for systematic errors to the quality of the standards. In general, the quality of the load element determines the quality of the measurement.

Finally, it is possible to use two probes with some VNAs to perform a calibration called "Thru-Line-Reflect" or TRL. This calibration uses the impedance of the line as a reference standard, and is most often used in automated wafer probing. This calibration properly calibrates the measurement for transmission and reflection.

All of the previous error correction methods depend upon the systematic errors being repeatable, and the standards being exceptionally good. It may be difficult to obtain a very good load standard for use on a PC board, and the mismatch in the probe may mask the measurement using the port extension technique. Further, placement errors in the probe with respect to the line can cause non-repeatable errors. Fortunately, the VNA has a mode which can be used to reduce the effect of these errors, namely, time domain transforms and gating.

In a VNA, the time domain transform is used to create a display of reflections as a function of time (or distance) down a transmission line. This is accomplished by using the inverse Fourier transform. The display may be configured to show the impedance as a function of distance down the line. The time domain transform can show the reflection from the probe-to-line connection, reflections from mismatches on the PC board line, and the reflection from the termination of the line, such as some active device.

The gating function of the time domain transforms allows one to eliminate the response from the measured data. In effect, the time response data is changed to show no reflection outside the gated region. Setting the gates properly can give the result of eliminating the reflection due to the probe connection, leaving only reflections from mismatches on the PC board. In this way, even relatively large mismatches caused by mispositioning the probe can be removed from a measurement, to give a truer picture of the return loss, match, or impedance of a device on a PC board. The capabilities and applications of time domain transforms are discussed more fully in the next section.

Time Domain Transforms:

The time domain transform is most useful used in the same manner as a traditional analog time domain reflectometer (TDR) and will be referred interchangeable with TDR in the remainder of this discussion. The TDR transform must be performed as a step response in the low pass mode to correspond directly to the analog TDR. In this mode, information about the phase of the reflections is retained, and the response shown is reflection coefficient vs. time (or distance). The distance down the line relates to the time by the propagation constant of the transmission line. The impedance of the line corresponds with the reflection coefficient; for a 50 ohm reference impedance, each percent reflection represents about 1 ohm change from 50 ohms. The exact equation is Z=Z0*((p+1)/(p-1)), a 0.05 reflection corresponds to 55.26 ohms, -0.1 corresponds to 40.81 ohms.

The TDR is very useful for identifying the impedance of sections of line, or for identifying the causes of mismatch in transmission line systems consisting of various types of lines and connectors. Without going in depth into the transform, the limitations of range (or length) and resolution depend upon the lowest .frequency and the bandwidth used in the frequency response measurement. The low pass mode requires that the frequencies used be harmonically related (i.e., evenly spaced), the minimum frequency is then set by the maximum frequency and the number of points chosen (Fmin=Fmax/(points-1)). The VNA will automatically adjust both the start and stop frequency to fulfill the last equation with the constraint of 1 Hz resolution on the start and stop frequency.

The TDR capability is ideally suited to make the RF probe most useful in investigating RF PC boards. Often, the impedance of lines on PC boards cannot be easily calculated, due to uncertainty in the board material or thickness, PC trace width changes with etching, and surrounding elements of the PC board and package. In these cases, the TDR is the easiest method for determining the characteristics of the PC board transmission line structures. The RF probe allows easy probing of various portions of a PC board. A limitation of the TDR is that it cannot resolve parallel transmission paths, so that the probe cannot be placed in the middle of a line, but must be placed at the end of a PC trace. Even so, the probe is useful in testing in the middle of circuits if some element, such as a coupling capacitor, can be removed to eliminate a second transmission path.

Figure 5 demonstrates the usefulness of the TDR by measuring the response of the RF probe when measuring a load on a 50 ohm line. This measurement was made on a 20 GHz network analyzer; the high bandwidth allows finer resolution in the time domain. The markers are placed at significant physical attributes: Marker 1 is set at the SMA connection of the probe. Marker 2 is set at the beginning of the spring contact exiting the probe. The flat trace between these two markers indicates a well matched line through the probe. Marker 3 is set at the peak of the TDR trace. This represents the impedance of the short length coplanar structure defined by the tip of the probe. The impedance of this section is about 57 ohms. Marker 4 represents the contact to the PC board and marker 5 shows the excess capacitance of the contact pad. In the next section, techniques are described which can remove these discontinuities of the probe from the return loss measurements being made down the line.

The TDR transform's usefulness is not limited to displaying impedance vs. distance, but can be used to enhance the frequency response measurements such as impedance or input match. As indicated earlier, a function known as gating may be used to remove known, unwanted reflections leaving only the unknown reflections from the test device. The gating may be applied with the analyzer in frequency response mode, as well as TDR mode. This gives the effect of an inverse transform on the gated TDR response to display the frequency response of the remaining reflections in the TDR trace. This is a very powerful tool for investigating PC board circuits even if it is difficult to make repeatable connections to the PC board.

However, as with all powerful tools, the quality of the results depends greatly on the setup and implementation of the TDR and gating functions. The TDR response and the gating effects are limited by the effective rise time of the system. This is set by the maximum frequency used in the VNA, and by another term known as the windowing factor. In short, the windowing factor determines the smoothness with which the endpoints of the frequency re-

sponse data are set to zero. Clearly, outside the frequency range of the VNA, the data is exactly zero. If the data in the transform is truncated, there will be "sidelobes" which appear as severe ringing on the TDR trace. Windowing has an effect of smoothly low pass filtering the data, which removes much of the ringing. Unfortunately, this has an effect similar to reducing the bandwidth, so the resolution will suffer. In viewing TDR responses, lowering the windowing rise time can aid in identifying and locating transitions in impedance, while increasing the windowing time can aid in smoothing the data to see the impedance value of different segments of the transmission line.

With gating, one must be aware that "looking through" a transition by gating it out results in a gated frequency response that is different from the true response, and generally has a lower reflection value (higher return loss) than actual. This change can be explained by understanding that each transition reflects some energy, so that farther reflections have less than 100 percent of the forward power available. When a large transition is gated out, the remaining reflections are not changed, thus some of the forward energy is effectively removed. For example, if a transition with a 0.1 reflection (10%, or 20 dB return loss, or SWR=1.21) is gated out, the remaining reflections are relative to .9, or will be 10% low. This would result in the return loss being about 0.92 dB low. Thus, it is very important that the RF probe used in conjunction with PC board testing have good performance to ensure accurate results after applying TDR gating enhancements.

Measurement Examples:

Figure 6 demonstrates the affect of gating on improving a measurement made with the probe. Figure 6A shows two measurements made with slightly different probe placements (approximately 1 mm). One measurement was made placing the probe to get the least ripple, the second was made by offsetting the probe slightly. The calibration for this measurement used an open/short/load cal with the cal kit modified to reflect the probe's parasitics as described earlier. The device under test (DUT) is a 25 ohm resistor at the end of a 43 mm long 50 ohm impedance line. The theoretical return loss should be about 9.5 dB plus about .3 dB/GHz due to line loss. The measurements differ by 2 dB up to 2 GHz, and by 3 dB to 3 GHz. The ripples become very large above 3 GHz, the error at 4.8 GHz is about 10 dB. Figure 6B shows the response when the portion of the TDR trace representing the probe tip is "gated out." The differences in the responses are less than 0.5 dB to 2 GHz, and less than .8 dB to 3 GHz. This represents a four times improvement in the repeatability of the measurement, with respect to probe placement, provided by the gating function. The response is much improved about 3 GHz as well.

Of course, it is most important that the measurement be correct, not just repeatable between probe placements. To attempt to provide some comparison, an identical DUT was constructed, but with a soldered-on planar SMA connector used as a launch. A TDR response of the launch enabled it to be tweaked to better than 40 dB to 3 GHz. A measurement of the DUT shows it to closely follow the ideal of about 10 dB return loss. Figure 7 shows a comparison of measurements of the 25 ohm resistor made with the soldered-on on connector and a similar DUT measured with the 3 GHz probe, with gating used to remove the probe transition. The results are remarkable in that the response is nearly identical up to 1 GHz. The results diverge by less than 0.7 dB up to 3 GHz. The fact that the gated response is somewhat lower in return loss is expected, and the 0.7 drop at 3 GHz represents 8% reduction in return loss. This is consistent with a 8% reflection for the probe transition or 22 dB return loss for that transition, which is what was measured for the probe back in Figure 4.

Finally, a calibration for a two port measurement using the TRL technique was performed on an HP 8720 Vector Network Analyzer. The calibration standards used were mating two probes directly for a "thru," using a short for the "reflect," and using line approximately 15 mm long for the "line." In this calibration technique, the exact length of line is not important. The impedance of the line sets the reference for reflection measurements. Figure 8 shows measurements of the 25 ohm DUT from figure 7, and a load at the end of a line. Here, gating was applied to remove the probe placement discontinuity. The results demonstrates very good measurements past 3 GHz. The line length chosen for the cal provides a calibration range of about 500 MHz to

4 GHz. The measurement of the 25 ohm DUT agrees remarkably well with the the earlier measurements in figure 7, especially considering two entirely different calibration techniques were used. A 10 dB return loss DUT makes a good test case as both return loss and source match affect the measurement.

Real World Applications for the K-50 probe.

Measurement is the basis of evaluation. In the realm of modern high frequency design, characterization measurements enable the reduction of theory to practice and the verification of theory from practice. In manufacturing, measurement enables ongoing control and optimization of product and process. Historically, the greatest obstacle to performing useful high frequency measurements has been accomplishing the needed connection between the measuring instrumentation and the device to be measured. The K-50 was developed to provide a universal means by which precision high frequency connections can be accomplished. The following actual situations exemplify the utility and capability of the K-50 in addressing the historical obstacles to accomplishing SIX SIGMA capable designs and manufacturability.

Using The K-50 To Perform Impedance Profile Measurements.

The essential reference circuit for a transistor is the supplier's line-fixture (production line test fixture). Transistor performance is specified in this circuit environment and compliance to specification can only be assured when the transistor is operated in an equivalent circuit environment. Prior to the development of the network analyzer and the K-50, circuit equivalency determinations were performed utilizing blind cut and try correlation methods, often with disastrous results. Figure 9 shows the actual Smith Chart Signature of a test circuit with an input matching network that is inversely correlated with respect to its line-fixture. That is, the measurement of transistor performance in an amplifier module test circuit does not correlate with the same measurement of the transistor in a line-fixture at the transistor manufacturing site.

A Smith Chart Signature is the visualizable function that is created when impedance profile information is plotted as a locus of points on a Smith Chart. The H.P. 8753 Network Analyzer measurement speed and Smith Chart display capability in conjunction with the universal connectivity of the K-50 Probe can be used to advantage by supporting the creation of equivalent circuits through real time tuning. In addition, the needed high frequency connections were typically soldered onto the linefixture commonly resulting in permanent performance degradation.

With the advent of the network analyzer having the ability to measure and display amplitude and phase data taken at numerous frequencies, and with the K-50's ability to provide nondestructive precision temporary high frequency circuit connections, the first absolute circuit equivalency determinations were successfully conducted. The equivalency referred to herein, means that the transistor views the application circuit and line-fixture circuit as having the same input and output matching network impedance profiles (the small signal impedance profile measurement methodology is depicted in Figure 10). An Impedance Profile is the tabular listing of measured impedance versus frequency values representing the transistor's view (measured with the transistor removed from the circuit) of its input and output matching networks taken with the normal circuit input and output ports terminated in 50 Ohms.

This measurement methodology was subsequently evolved into an R.F. design methodology and was first applied to problem solve an R.F. power amplifier circuit that was exhibiting substantial batch performance variation. Whereas all previous attempts to understand and resolve this amplifier performance problem had met with failure, the new design methodology readily revealed the problem. The amplifier had a rotated input matching network impedance profile relative to that of the transistor line-fixture (shown in Figure 11). Even laser tuning could not accommodate transistors that required an input matching network impedance profile that deviated from the intersection of the profiles shown in Figure 11 to any significant extent. For proper performance, the test fixture and the amplifier module should provide the same impedance profile, then the transistor manufacturer's testing will ensure all amplifier modules will meet specifications.

Next the fledgling R.F. design methodology was applied to the development of a new R.F. power amplifier circuit. The amplifier input matching network design was determined based on the measured impedance profile of the corresponding transistor line-fixture (refer to Figure 12 for the input matching network impedance profiles, the output matching network impedance profile was known from previous design work and is not shown in this example). As a test of the design methodology, 50 transistors were selected for optimum performance in the supplier's line-fixture and sent to the customer. At the customer site a test fixture was made by tuning the impedance profiles of the input and output matching networks to be the same as those of the line-fixture (this tuning was performed in real time using the H.P. 8753 Network Analyzer in the Smith Chart display mode of operation). When the 50 transistors were measured in the customer's real time correlated test fixture, all 50 yielded identical measured performance. When 100 of the new amplifier circuits were fabricated in the customers factory, all 100 passed their test specifications with substantial margin. Supported by this design methodology, the design task was completed in six weeks as opposed to prior efforts that required six months and did not accomplish defect free manufacturability. With the need to substantially reduce design cycle time in order to remain competitive, this design methodology establishes an essential capability.

Whereas R.F. designers once attempted to produce the needed circuit equivalency by making line-fixtures and application circuits physically similar, practical network analyzer real time impedance profile measurement capability has allowed physically dissimilar circuits to become explicit equivalents of each other through tuning (a transistor line-fixture is shown in Figure 13 and its equivalent application circuit is shown in Figure 14). This is the basis of SIX SIGMA R.F. performance compliance by design.

The K-50 In Support Of WORLD CLASS Manufacturing Capability.

The K-50 is especially well suited to manufacturing applications where its precision, consistency, reliability and ability to effect a highly repeatable universal high frequency connection are considered valuable attributes. The K-50 is totally automatable and is highly utilized in the manufacture of connectorless, high circuit density, high frequency products that employ surface mount component technology. Its universal open tip architecture does not require a special contact target thus promoting usage as a common connection device across product lines throughout a factory. For example, Motorola utilizes the K-50 to manufacture:

> Portable Cellular Telephones (refer to Figures 15 through 17),

> Mobile Cellular Telephones,

> Cordless Telephones,

In these applications the K-50 has enabled:

> total manufacturing automation,

> the establishment of common fixtures for multiple products,

> the building of multiple products on the same production line,

> the elimination of fixture change-over down time,

> the elimination of ongoing fixture correlation as required by tuned R.F. connection schemes,

> production personnel to maintain their fixtures by performing pin changing,

> and award winning measurement consistency.

A typical automated test application of the K-50 involves operating it at 90% of full mechanical compression and terminating its SMA connector in a 6 dB coaxial attenuator. By way of example, when operated in this manner 900 MHz. 12 dB SINAD measured at a -115 dBm signal level will typically exhibit only 0.2 dB variation.

The K-50 As A Design And Manufacturing Tool.

Perhaps the most important aspect of the K-50 is its applicability as a common measurement tool in support of Design For Manufacturability. When employed as a common measurement tool, Research, Development, Factory, and Field Service personnel can all perform and exchange measurement data in a directly comparable format. With Hewlett-Packard supporting error corrected measurement, design calculations can be directly related to measured values performed by any of the technical functions

or groups within a corporation. The application possibilities and potential benefits are virtually limitless.

Motorola has elected to contribute the K-50 technology for the common benefit of the electronics industry (digital technology included). The K-50 could have accomplished only a limited usefulness if held back proprietarily, but if openly shared to provide the industry with a common basis for performing high frequency measurements has the potential to be profoundly beneficial. Unique to this type of invention, it is not possible for one party to derive a substantial benefit to the exclusion of others. For the SIX SIGMA capable design methodology to succeed, an industry majority must accept and use the K-50 tool. Your participation and contributions are, therefore, needed for the common good of our industry. With your support we can all be on the road to SIX SIGMA capable high frequency designs.

Conclusions:

In this paper we have presented a new concept for probing RF circuits without the use of solder on connectors. The K-50 probe has demonstrated usefulness in both the development and manufacturing arenas. It provides a high performance contact to RF circuits, and enables careful characterization of PC Board components in a simple and economical manner. Additionally, the probe is ideally suited for automated board test applications where reliable RF measurements are needed. The calibration and time domain capabilities of Vector Network Analyzers enable very accurate measurements to be obtained even considering difficulties such as probe placement and contact repeatability.

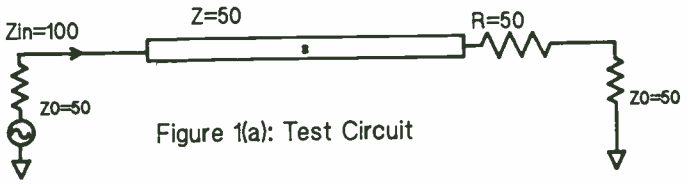


Figure 1(a): Test Circuit

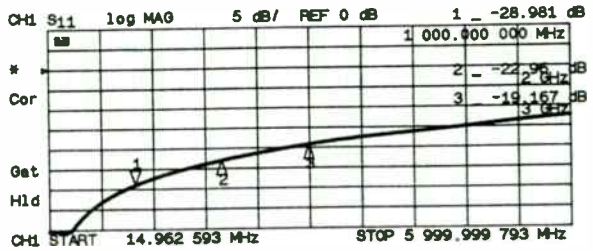


Figure 1(b): Return Loss of Test System Interface

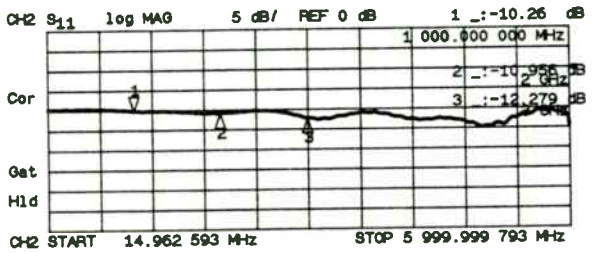


Figure 1(c): Return Loss of Test Circuit by Itself

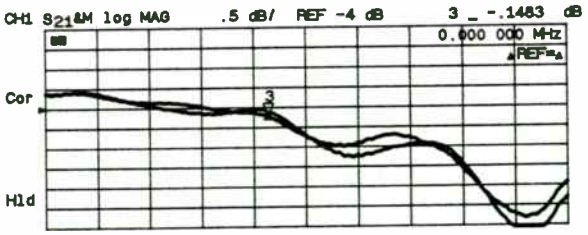


Figure 1(d): Gain Variation Due to Interface

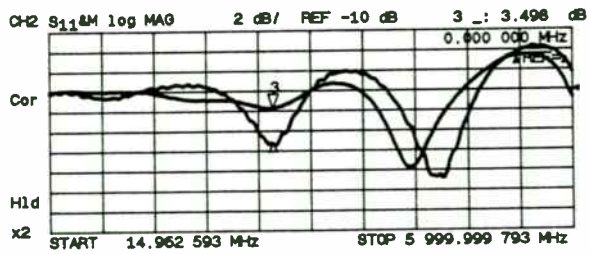


Figure 1(e): Variation in Match Measurement Due to Interface

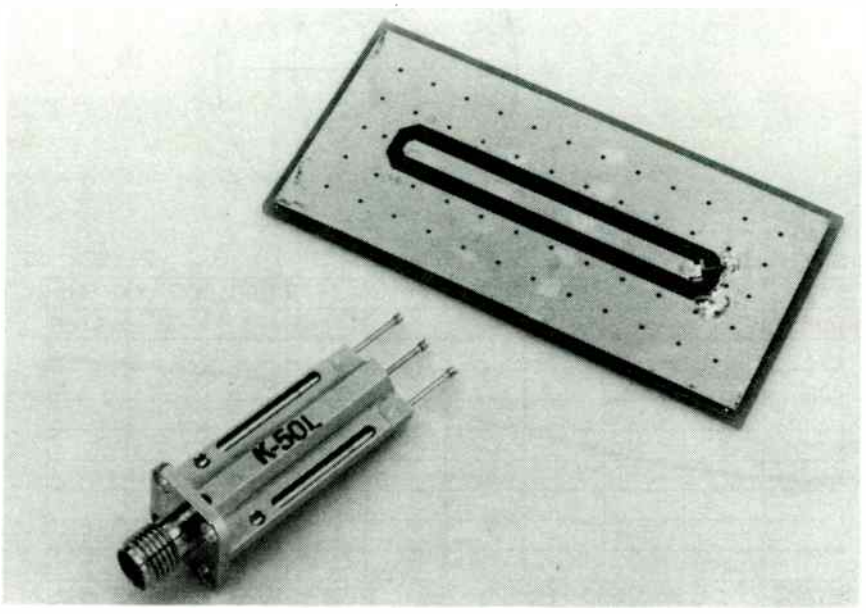


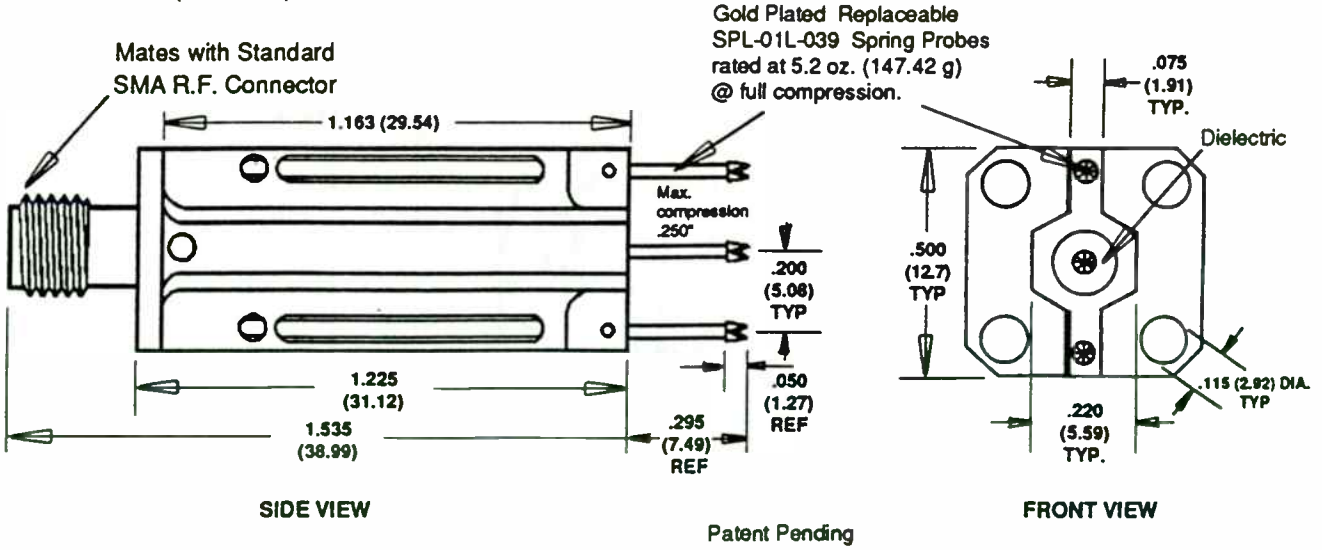
Figure 2: The K-50 Probe and Test PC Board

EVERETT CHARLES TECHNOLOGIES
Contact Products Division

Broadband 50 Ohm Coaxial Test and Measurement Probe

Model Number: K-50L

Dimensions in inches (millimeters)



Patent Pending

Figure 3: Mechanical Details of the K-50 Probe

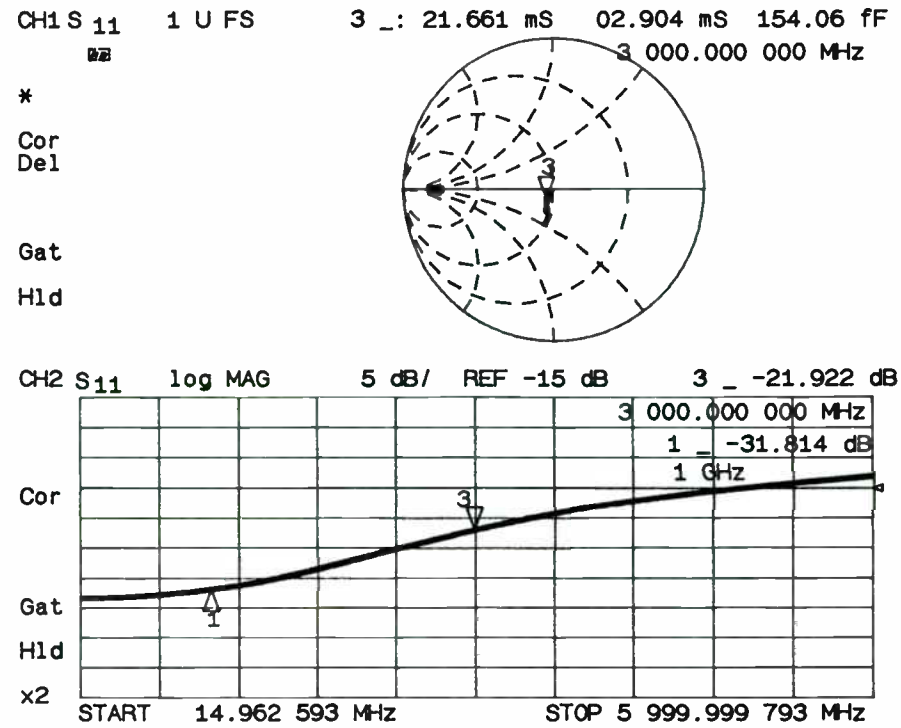


FIGURE 4: The Smith Chart shows the probe impedance, with 0.15 pF cap. The rectangular plot is the return loss of the probe in dB.

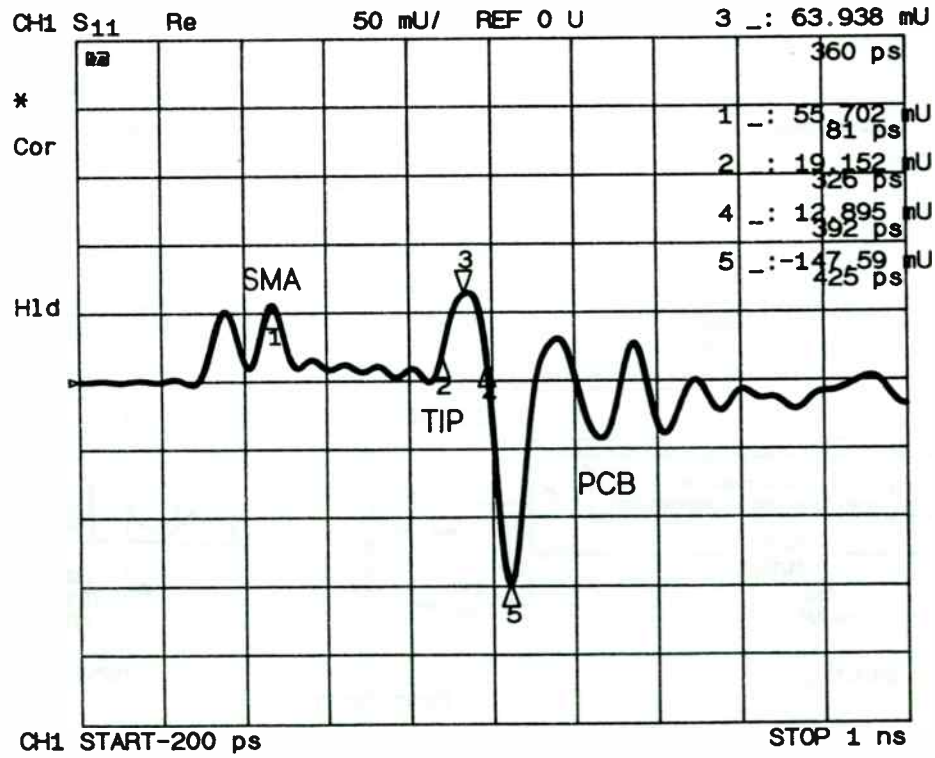


Figure 5: TDR response of the K-50 Probe

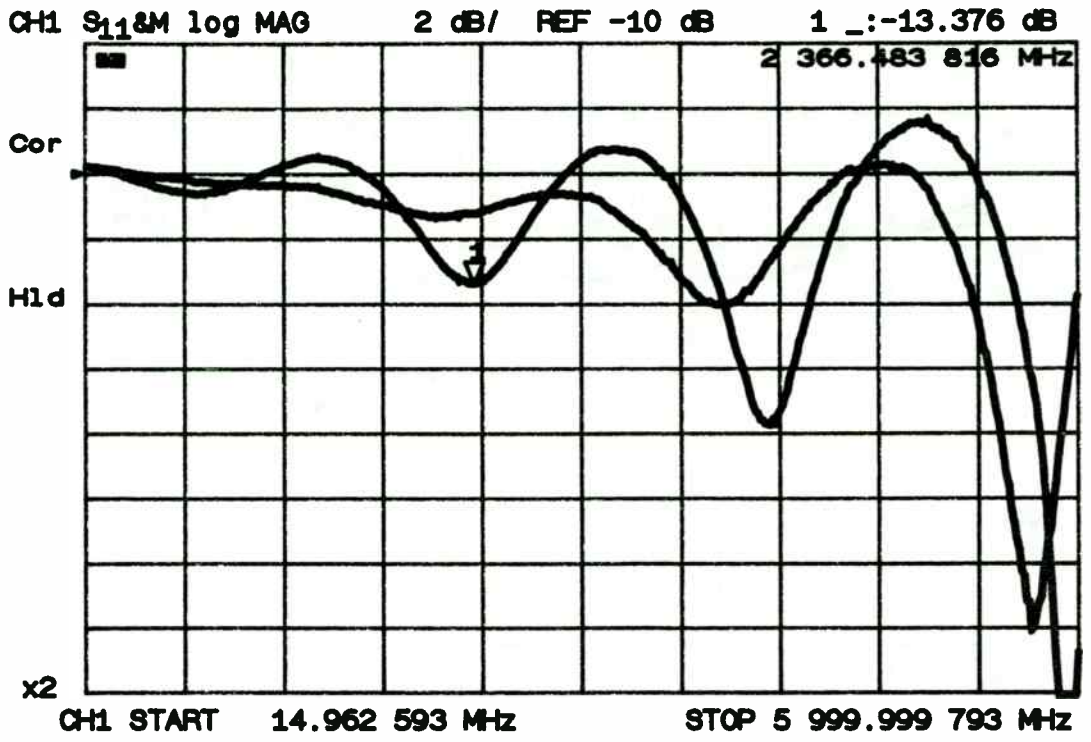


Figure 6A: Return Loss with two different probe positions.

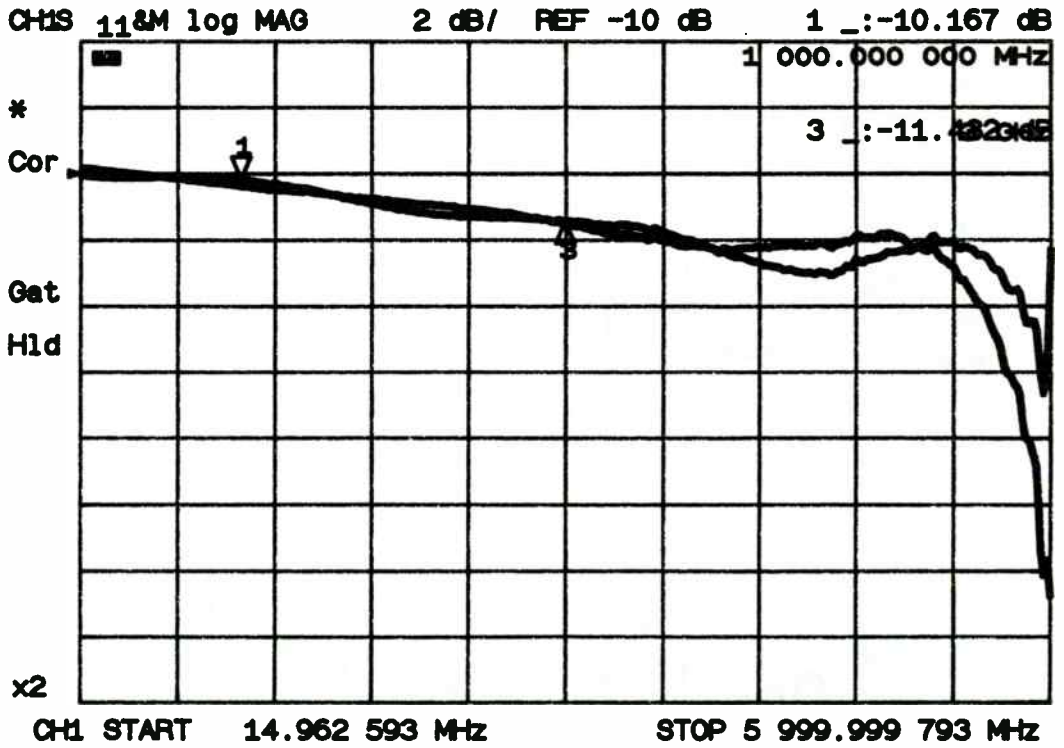


Figure 6B: Same measurement with TDR gating applied to the frequency response

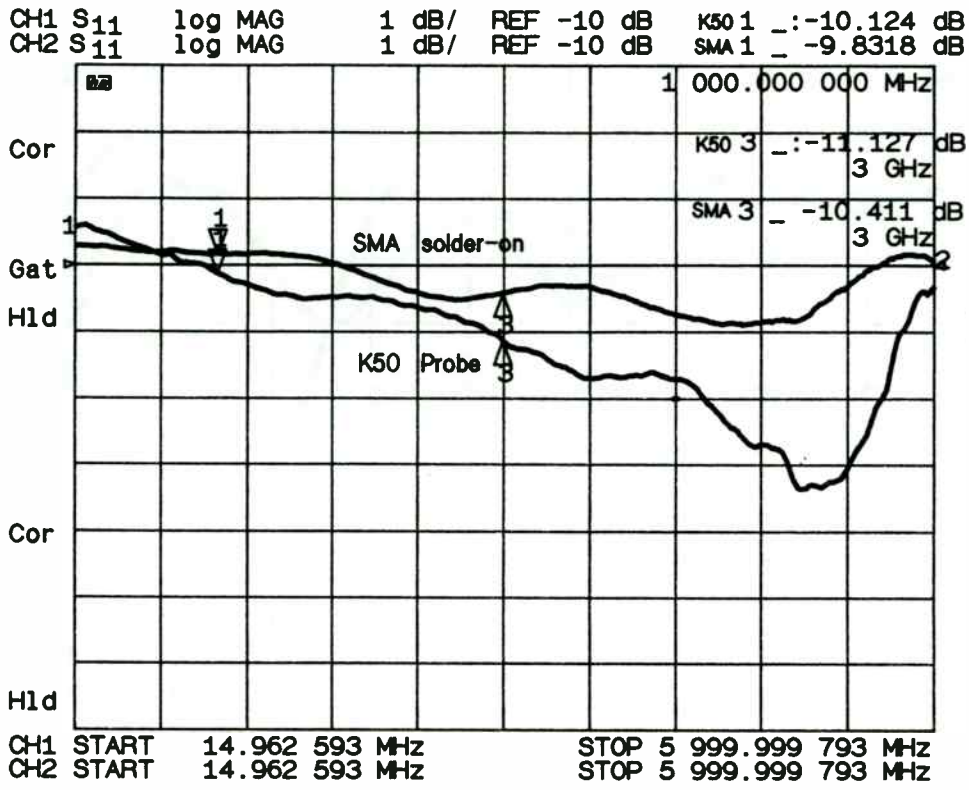


Figure 7: Comparison of SMA connection with K50 Probe (and gating).

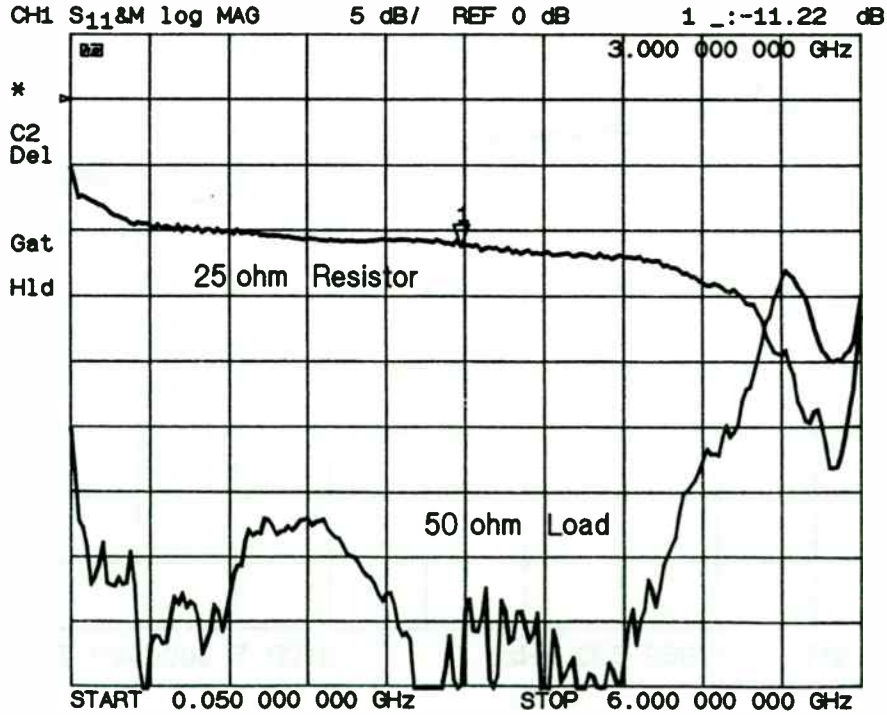


Figure 8: Measurement of a 25 ohm resisitor, and a 50 ohm load with a TRL calibration of the probe, and TDR gating

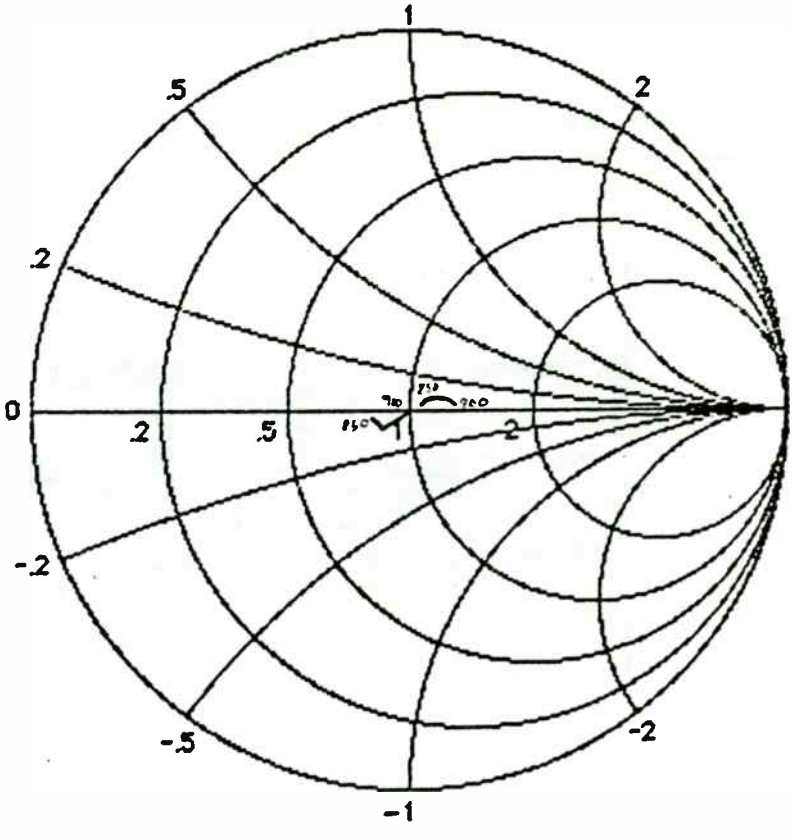


Figure 9: Smith Chart Signatures -
Test Circuit and Line-Fixture

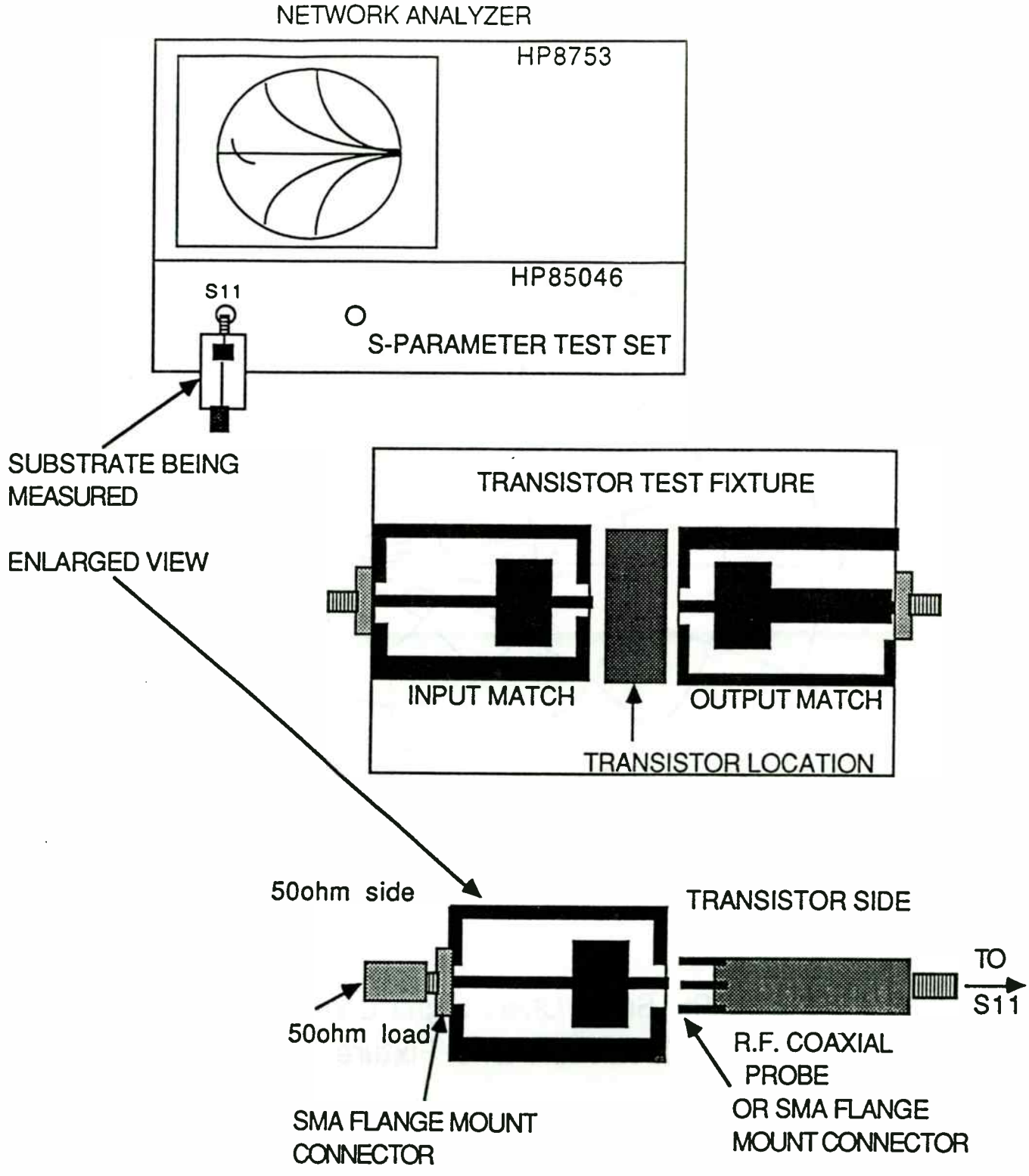


Figure 10: Impdedance Profile Measurement Methodology

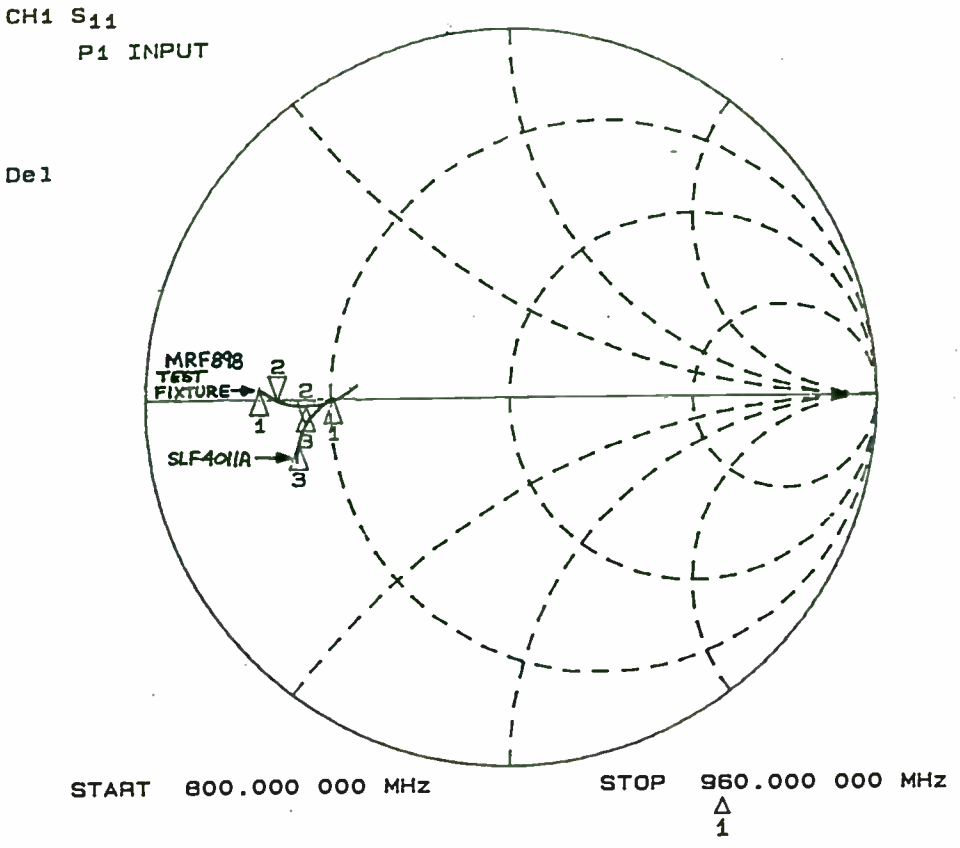


Figure 11(a): Impedance Profiles Relative to Line Fixture

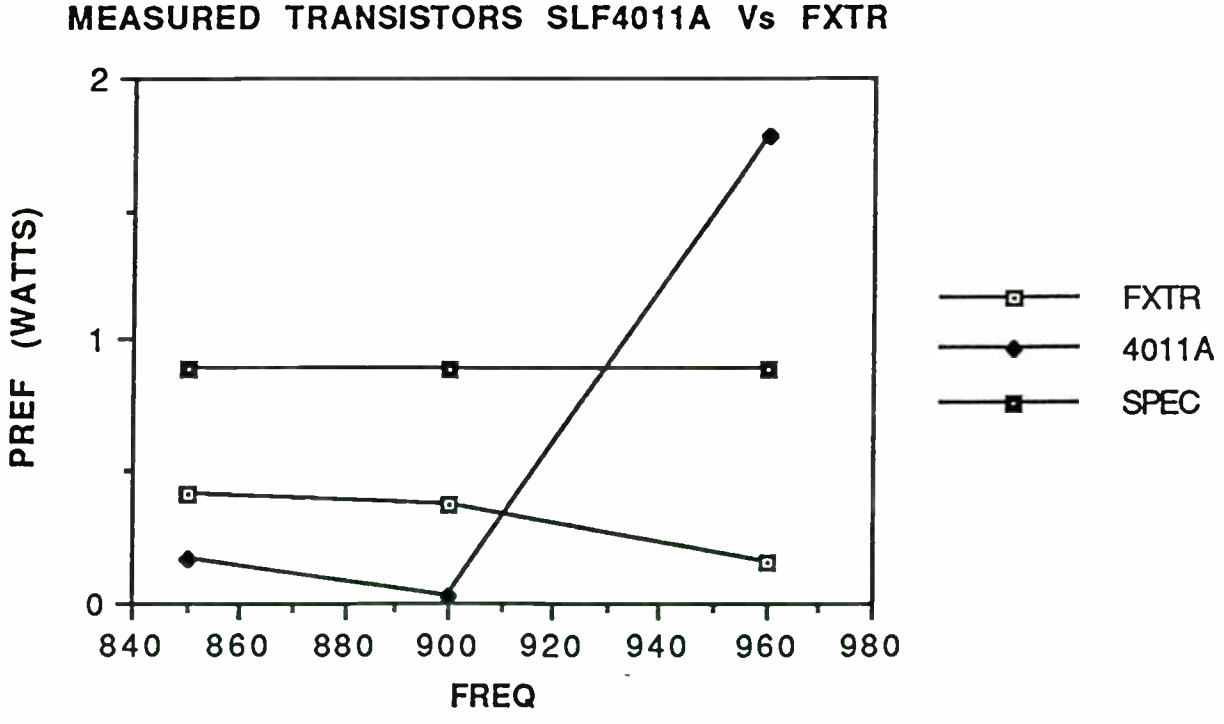


Figure 11(b): Test Results Showing Out of Spec. Performance

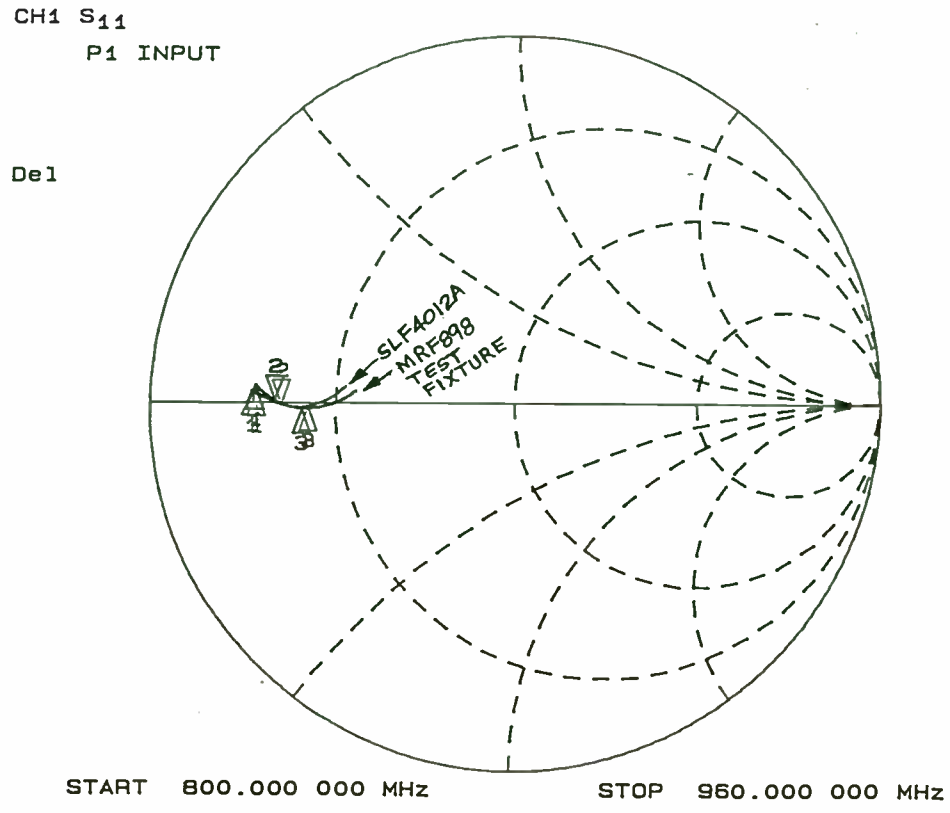


Figure 12(a): Impedance Profiles – New Methodology Design

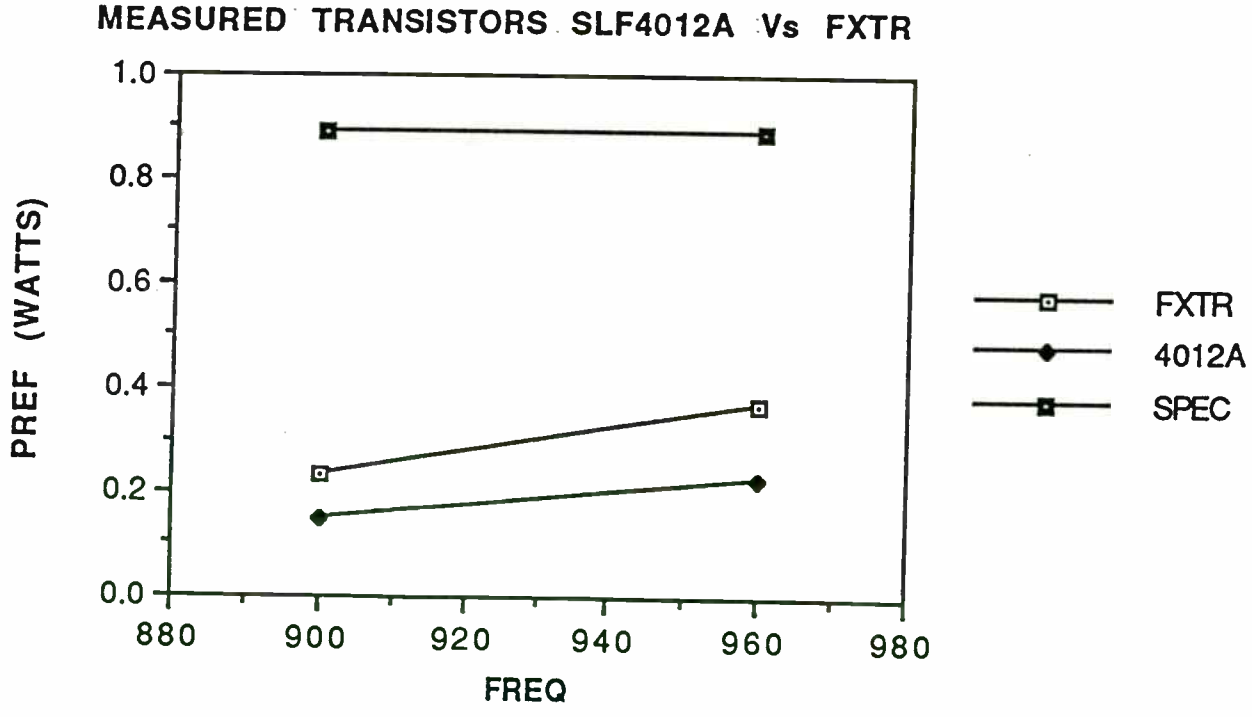


Figure 12(b): Test Results Showing Margin on Spec'd. Perf.

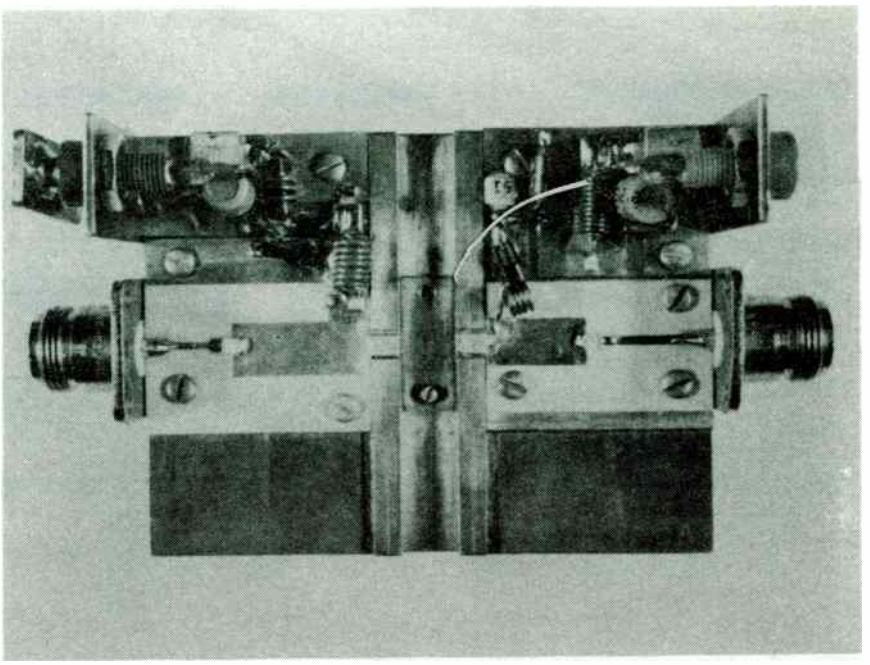


Figure 13: Transistor Line Fixture

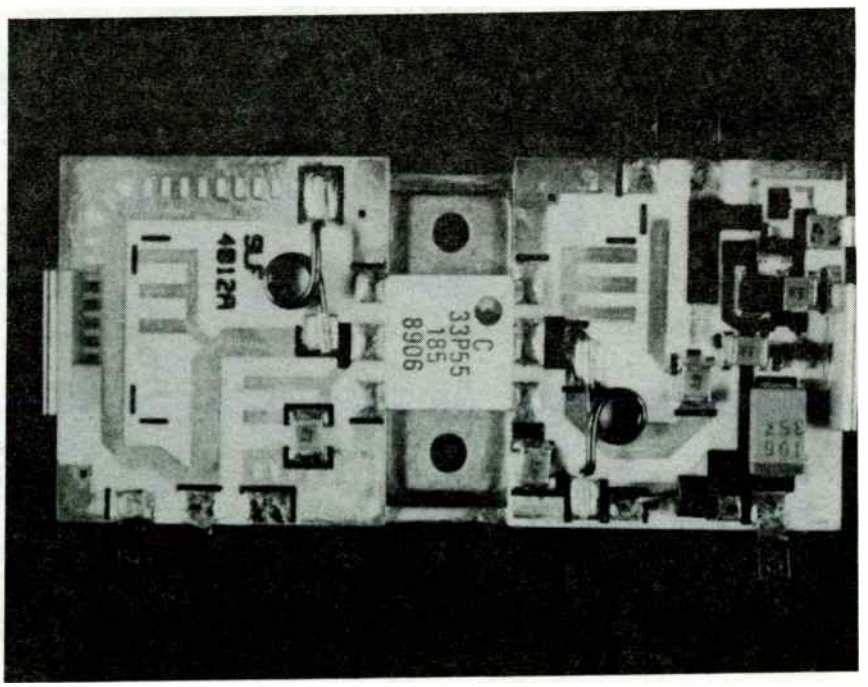


Figure 14: Application Circuit

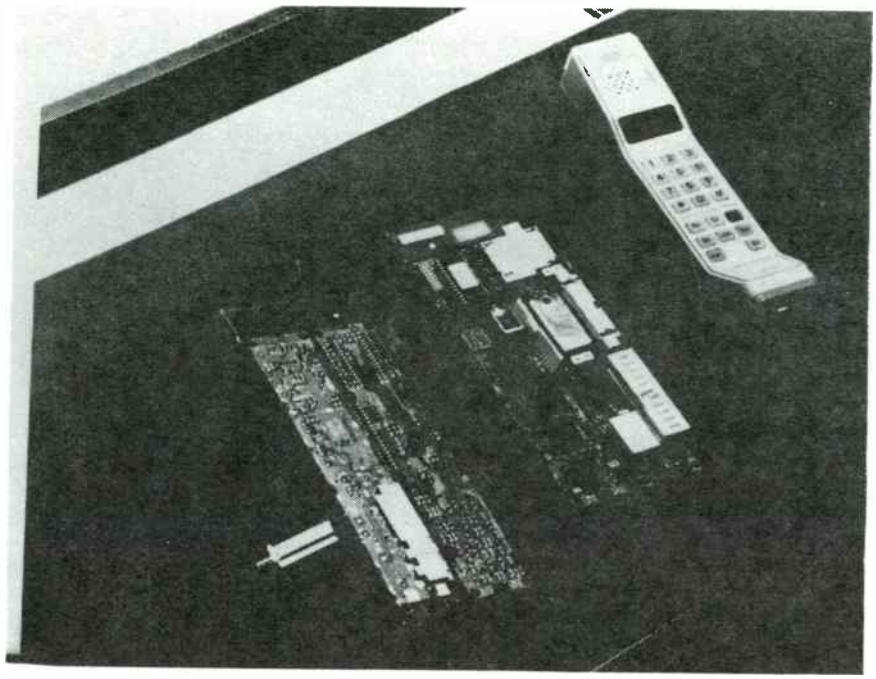


Figure 15: A K-50 Probe with A Cellular Telephone Board

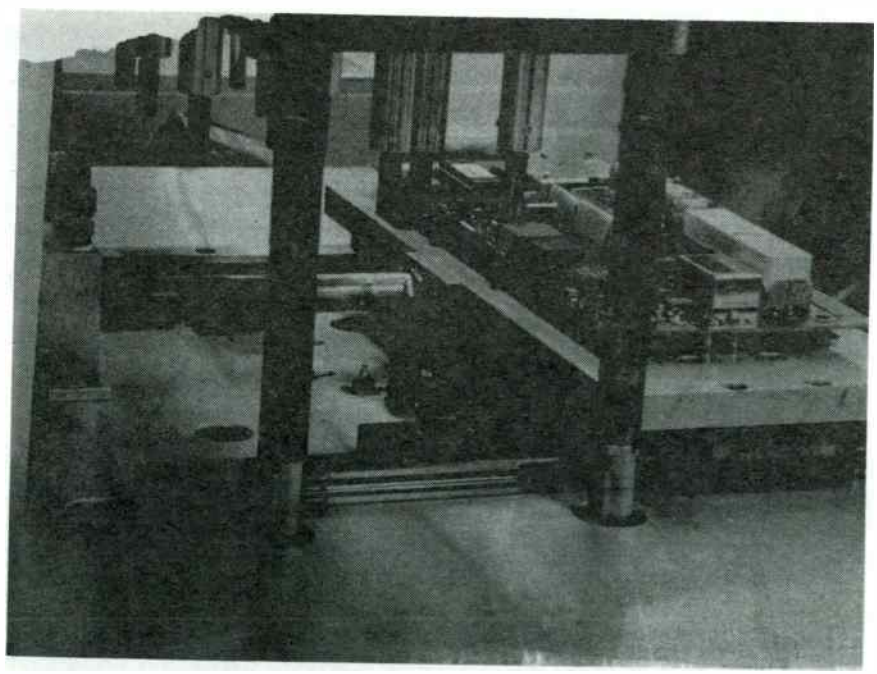


Figure 16: The K-50 Probe in an Automated Board Tester

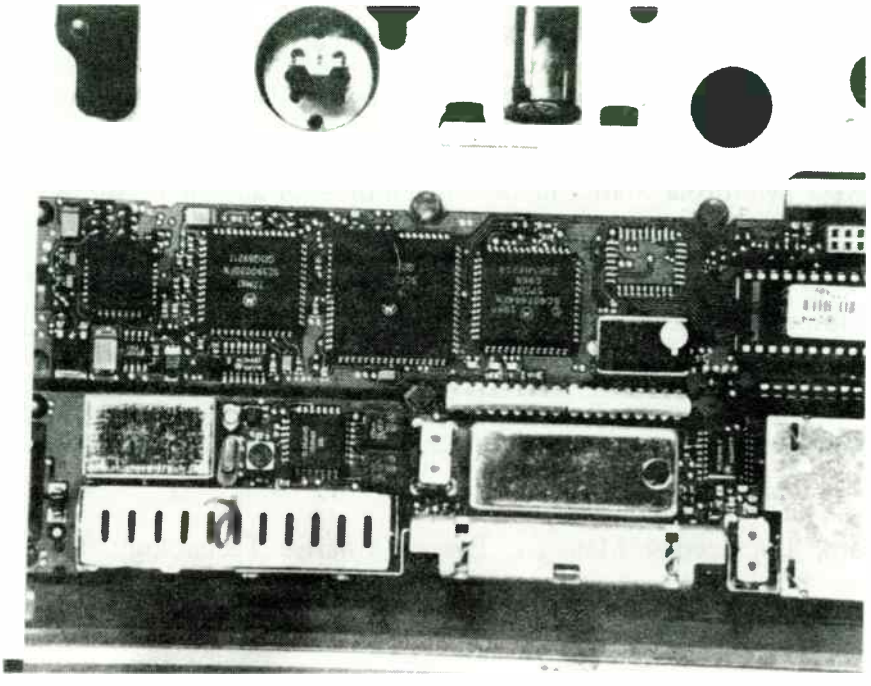


Figure 17: A K-50 Probe in an Automated Test Fixture

Contributors:

Joel Dunsmore - Figures 1,2,4-8

Bob Kornowski - Figures 13-17

Jim Long - Figure 9

Miles Tusa - Figures 10-12

Chuck Tygard - Figure 3

Acknowledgments:

Joel Dunsmore, R&D Development Engineer, Hewlett-Packard, Microwave Instruments Division.

Bob Kornowski, Motorola Staff Engineer/Inventor - Originator of the K-50 Probe, the SIX SIGMA High Frequency Design Methodology, and the "QUALITY through INVENTION" Continuous Quality Improvement Culture.

Jim Long, Motorola Principle Staff Engineer - An early investigator and proponent of using the network analyzer as a problem solving tool.

Miles Tusa, Motorola Lead Engineer - Miles was the first to apply the SIX SIGMA High Frequency Design Methodology, involving Impedance Profiles and Smith Chart Signatures, to develop a new radio product.

Chuck Tygard, Engineering Manager, Everett Charles Technology, Contact Products Division.

Low Cost RF Tuner System for JDC Load Pull and SSPA Design

Christos Tsironis and Dominique Dubouil
Focus Microwaves Inc.
277 Lakeshore Road
Pointe-Claire, Quebec
Canada H9S 4L2

Summary
Focus Microwaves presents a Load Pull and Noise Measurement and Design System, conceived for the Basic Needs of RF Design and Test Engineers:
-Accuracy, -Low Cost, -Focus on Key Applications in Power and Low Noise and -Design Capability for Power Amplifier stages.
The new system includes all hard and software components required to configure a Test and Design Workstation based on a 80-386 or -486 IBM-PC and most commonly used GPIB equipment.

Introduction
The important market of Personal Communications (PCN) in the low microwave and RF area of 800 to 3000 MHz requires:
- New low cost, reliable and state of the art active devices (FETs, MOSFETs, bipolar transistors...)
- Efficient designs to minimize the requirements to the devices for given system performance and
- Sophisticated test methods to detect problems early in the Design Cycle.

It became common knowledge, meanwhile, that automatic load pull and noise measurement systems provide great help in understanding the devices and speed up the test and design process.
Some of the new amplifier designs are based on data generated by those computer controlled tuner systems, mostly for low noise applications, but progressively also for high power.
Still there is some time to go, before those systems develop their full potential and become as reliable, understandable and User friendly as most Users would like to have. Nevertheless the potential is there and spreading those systems in as many labs as possible will help their evolution.

Price has always been a factor keeping computerized tuner systems out of most labs. At 60 to 90,000 $ most managers require a thorough justification to approve, but this last one only comes with broad use. So we have a kind of self-blockade of the cycle.

We have developed the **Microwave Tuner System (MTS)** with a close look to the core needs of the RF Design and Test Engineers in the Product Development and Production teams:
- **Accuracy**: The MTS tuners cover the frequency range of 800 MHz to 3 GHz and provide state of the art 50 dB RF-impedance resetability (±0.003 reflection factor units).
- **Low Cost**: The MTS costs about 1/3 of other (more sophisticated) computerized tuner systems.
- **Key Applications**: The MTS measures all important RF quantities; ie. Power, Gain, Efficiency, Intermod as well as 4 Noise parameters.
- **Design**: The MTS permits to transfer ISO Power contours to a Network Simulator and optimize High Power Amplifier stages.

The MTS Components and Capabilities
The MTS includes the following components:
- Two computerized tuners, model MTS-308 in SMA, GPC-7 or N-connector configuration
- One PC insertable tuner controller
- GPIB interface
- Setup, test fixture and tuner calibration software
- Measurement software
- ISO contour generation graphics software
- RF power amplifier design software (optional)
- Preselectable GPIB drivers for over 50 popular instruments (network analyzers, power meters, spectrum analyzers, dc bias controllers, signal sources, frequency counters and noise analyzers).

Figure 1 shows a typical setup for Load Pull and Noise Measurement.

The MTS permits the following operations:
- **Calibration** of the tuners, the setup components

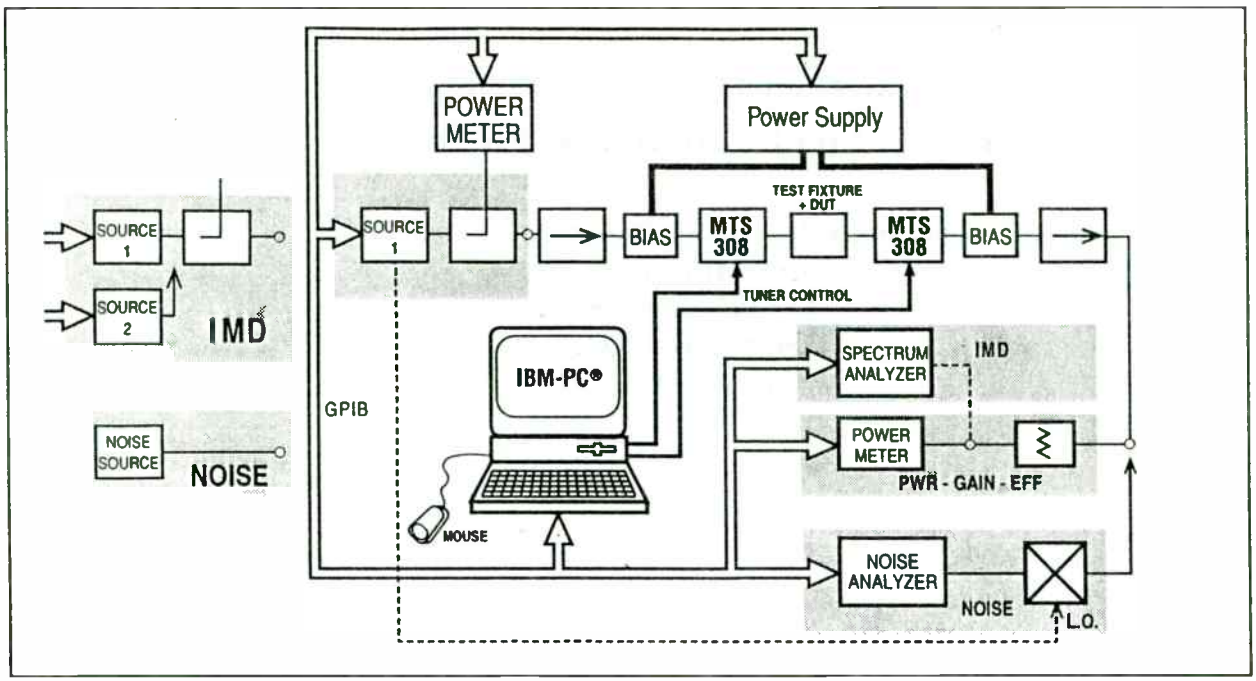


Figure 1. The MTS Load Pull and Noise Measurement setup

and test fixture using any of the available network analyzers. In particular the test fixture is characterized using generic TRL formulas, which are valid for any kind of fixture.

- **De-Embedding** to any reference plane selected by the User.
- **Mouse tuning** to any point on the Smith Chart (not only calibration points).
- Measurements at **selected** Smith Chart **areas**, in order to avoid oscillations.
- Classical **Load/Source Pull** of Power, Gain, Intermod and Efficiency. Further options include JDC (Japanese Digital Communication) or adjacent channel leakage tests, high order intermod and oscillator load pull tests.
- **Automatic search for best performance** in power, gain, efficiency both at the source and load side of the device.
- **Saturation measurements** of power, gain and efficiency.
- Automatic search for **optimum Noise** match and **4 Noise parameters**

MTS-308 Tuners

Figure 2 shows the MTS-308 tuners with SMA connectors. The MTS-308 are mechanical tuners using a parallel slotted airline as transmission media and a metallic RF probe to generate controllable reflection. When the probe is withdrawn the tuner behaves like a transmission line, thus avoiding parasitic oscillations. When the probe is inserted the tuner has always a low pass behaviour below ≈700 MHz, with the same effect. It is obvious that among all types of variable tuners the slotted line type ones are the best compromise for parasitic oscillations.

The probe is moved using two stepper motors and a rugged translation mechanism. Most of the MTS-308 parts are 'of the shelf' articles and thus permitted low manufacturing cost and high reliability. Size and weight of the MTS-308 have been optimized to permit (manual) operation down to 750 MHz and up to 4.2 GHz (optional). The reduced weight permits to position MTS-308 tuners on wafer probe stations very easily and without any implications due to vibrations.

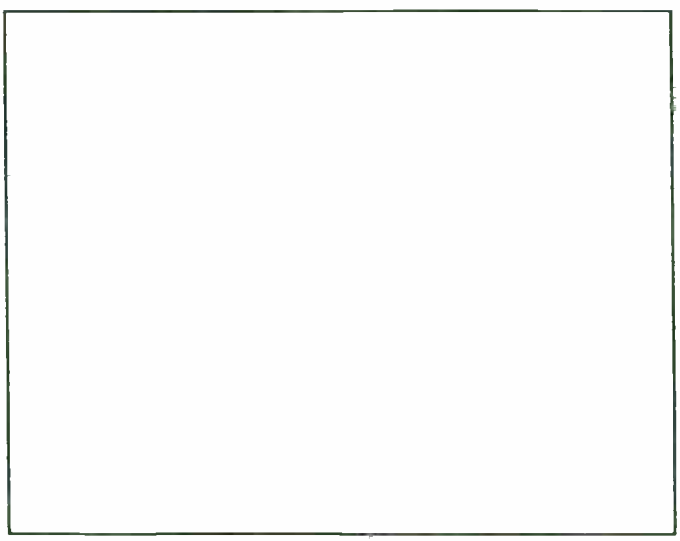


Figure 2. The MTS-308 tuners with SMA connectors

The MTS-308 Specifications

A key advantage of the MTS tuners is that they can handle practically illimited RF power, either in CW or pulsed form. Because of the wideband behaviour pulsed operation is not a problem. Also fine tuning to millions of impedance states is a key benefit. The MTS-308 specifications are listed in Table 1. Figure 3 shows the MTS-308 tuning capability.

- Frequency range: 800 to 3000 MHz (4200 MHz optional).
- VSWR min: 1.12:1 (SMA), 1.06:1 (GPC-7).
- VSWR max: 10:1 min (SMA), 12:1 min (GPC-7), Typical 15:1.
- Tuning Resolution: 0.18° per step at 3 GHz.
- Insertion Loss: 0.4 dB (SMA), 0.15 dB (GPC-7).
- RF Resetability: > 50 dB.
- Total size: 300 x 150 x 175 [mm] or 11.8 x 5.9 x 6.9 [inches].
- Weight: 4.1 kg.
- Tuning Speed: 360° tuning in 1000/f[MHz] x 15 seconds.
- DC power requirements: 12V, 3A max (provided by PC controller).
- Humidity, Temperature: normal laboratory conditions.
- Vibrations: The MTS-308 units are not sensitive to moderate shocks and vibrations both from operation and accuracy points of view.
- Validity of calibration data: 1 to 3 months of normal operation.
- Calibrated points (Load Pull): 181.

Table 1. MTS-308 Specifications

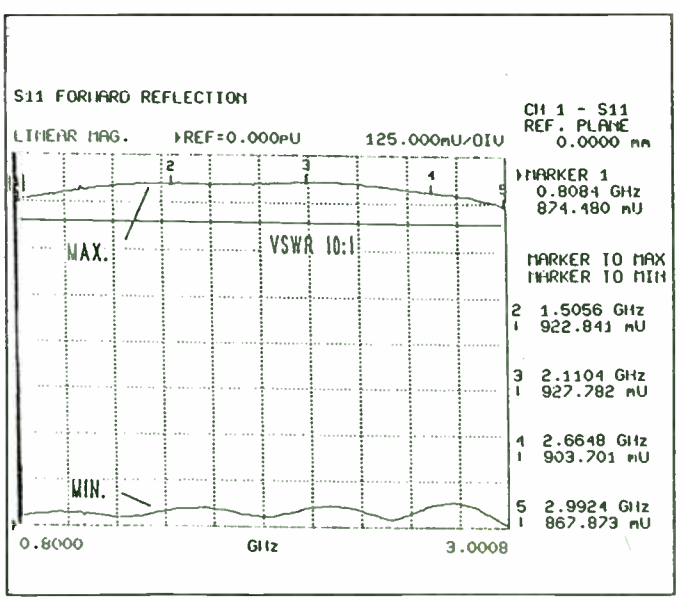


Figure 3. MTS-308 tuning capability

Table 2 shows S-parameter dispersion when the tuner is moved 10 times to the same set of positions.

RF-Resetability over 10 Cycles (0.001->60dB, 0.0001->80dB)

TUNER MTS#02, Frequency = 0.800 GHz

Point	δ|S11|	δφ11°	δ|S12|	δφ12°	δ|S21|	δφ21°	δ|S22|	δφ22°
1:	-0.0014,	0.02	0.0007,	0.06	0.0006,	0.06	-0.0013,	0.11
2:	0.0006,	-0.04	-0.0004,	-0.04	-0.0004,	-0.04	0.0005,	-0.03
3:	0.0019,	-0.17	-0.0017,	-0.12	-0.0017,	-0.13	0.0020,	-0.06
4:	-0.0018,	0.14	0.0015,	0.12	0.0015,	0.12	-0.0020,	0.08
5:	-0.0007,	0.04	0.0005,	0.05	0.0005,	0.06	-0.0007,	0.06
6:	0.0001,	-0.03	-0.0001,	-0.01	-0.0001,	-0.01	0.0001,	0.01
7:	0.0006,	-0.03	-0.0005,	-0.03	-0.0005,	-0.03	0.0005,	-0.03
8:	-0.0005,	0.03	0.0008,	0.04	0.0008,	0.05	-0.0005,	0.05
9:	-0.0003,	0.04	0.0004,	0.04	0.0004,	0.04	-0.0003,	0.03
10:	-0.0016,	0.10	0.0017,	0.11	0.0018,	0.11	-0.0014,	0.12
11:	0.0010,	-0.08	-0.0016,	-0.10	-0.0015,	-0.11	0.0012,	-0.12
12:	-0.0020,	0.06	0.0020,	0.11	0.0020,	0.11	-0.0017,	0.16
STD.DEV=	0.0012,	0.08	0.0012,	0.08	0.0012,	0.08	0.0012,	0.08

'S' parameters of above Points ...

1:	0.390,	138.1	0.897,	56.8	0.898,	56.7	0.383,	157.6
2:	0.586,	41.0	0.777,	48.1	0.777,	48.0	0.593,	-124.2
3:	0.620,	-40.8	0.755,	48.3	0.755,	48.3	0.625,	-42.9
4:	0.607,	-85.1	0.767,	48.6	0.767,	48.6	0.607,	2.1
5:	0.574,	-123.9	0.790,	49.2	0.790,	49.2	0.575,	42.7
6:	0.519,	166.0	0.822,	49.9	0.823,	49.8	0.529,	115.3
7:	0.715,	152.9	0.652,	38.4	0.653,	38.3	0.731,	104.6
8:	0.772,	-106.2	0.600,	38.2	0.601,	38.2	0.776,	2.4
9:	0.782,	-51.2	0.590,	38.4	0.590,	38.4	0.787,	-52.4
10:	0.770,	-7.8	0.601,	38.4	0.601,	38.4	0.777,	-95.5
11:	0.780,	66.1	0.582,	34.2	0.582,	34.1	0.777,	-176.9
12:	0.737,	113.5	0.639,	37.1	0.640,	36.9	0.724,	141.6

Table 2. RF resetability

The MTS Operation

The MTS operation consists of 3 steps: **1.-Calibration, 2.-Measurement and 3.-Data Processing**. In case of **Amplifier Design** this is then a 4th step.

Calibration

All passive components of the Load Pull (or Noise) measurement system have to be pre-characterized (calibrated) using an automatic network analyzer. This includes the tuners (once every couple of months), isolators, bias tees, cables, attenuators and test fixture (calibration only once).

Test fixture calibration consists of using TRL standards to generate S-parameters of both fixture's halves using MTS's software. All calibration data are saved on harddisk files and are reusable at any time or transferrable to another computer. This allows maximum flexibility and mobility of the system.

Measurement

Once the DUT is inserted in the test fixture, the MTS software controls bias and signal source power and frequency to perform a multitude of automatic or manual measurements.

The **Setup** data are loaded automatically and the User has the choice of different reference planes to perform the measurement.

This makes it easy to evaluate the effect of different test fixtures on the device's performance or either study the effect of different device packages.

The tuners are **moved manually** either using the PC's cursors or by direct tuning with the mouse and phase and amplitude corrected measurements are made via the GPIB connected and preselected instruments. The results of manual measurements are saved on ASCII (printable) files on the harddisk.

In **automatic measurements** the User can preselect a number of key settings and parameters and to perform either automatic search for an optimum (minimum Noise Figure or maximum Power for example). Or he can scan the complete Smith Chart with the tuners and generate binary Load Pull Data files which are used by the graphics software to generate ISO Contours (of Power, Gain, Efficiency or Intermod).

In case of **risk of oscillations** the User has the option to generate an **Impedance Pattern** on the Smith Chart using mouse tuning, save it on a harddisk file, and measure only along the points of this pattern. This pattern can be retrieved voluntarily, modified and resaved. Since the points are saved in Impedance (and not in tuner position) form this pattern file will generate the same impedance setup at all frequencies, independently of the actual calibration. This is important and deliberates the User from the worry to know before the tuner calibration which points he will need to measure later on.

Many power (and low noise) transistors have **very low input impedance** in the order of 1 to 2 Ω. Whereas the optimum noise reflection factor can be computed from measurements in other areas of the Smith Chart (due to the linearity of noise behaviour), in the case of power load (or source) pull the device has to be really presented the required impedance in order to show its behaviour. If a variable tuner generates such low impedances, ie. almost a short circuit, then the overall measurement accuracy will be unacceptable, this including the calibration accuracy of the automatic network analyzer.
The simplest and safest way around this measurement problem is the use of microstrip transformers. The simplest transformers are $\lambda/4$ sections of microstrip line which permit load impedances down to 0.5 Ω very easily. The bandwidth is reduced indeed, but this can be cured by using multisectional transformers. Frequency ratios of as much as 3:1 can be covered using 4 sections. It is important to realize that even the smallest transformed impedance is not exactly the same as the Zin* the tuner will find it in the area any how. Using an ordinary single section microstrip transforming network we were able to generate impedances as low as 0.7 Ω using the MTS-308 (figure 4) with very good measurement accuracy.

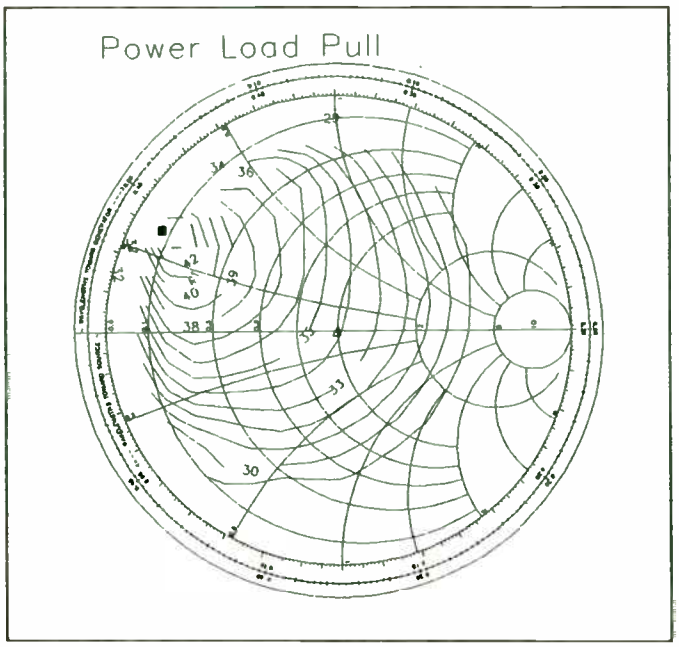


Figure 4. Load Pull using $\lambda/4$ transformers, $\Gamma_{opt} \approx 0.9$

Data Processing
The load pull files and the saturation (power transfer) files generated by the MTS software can be processed to plots:
- The load pull data files to ISO Power, Gain, Efficiency or Intermod contours (figures 5,6)
- The power transfer files to XY-plots (figure 7)

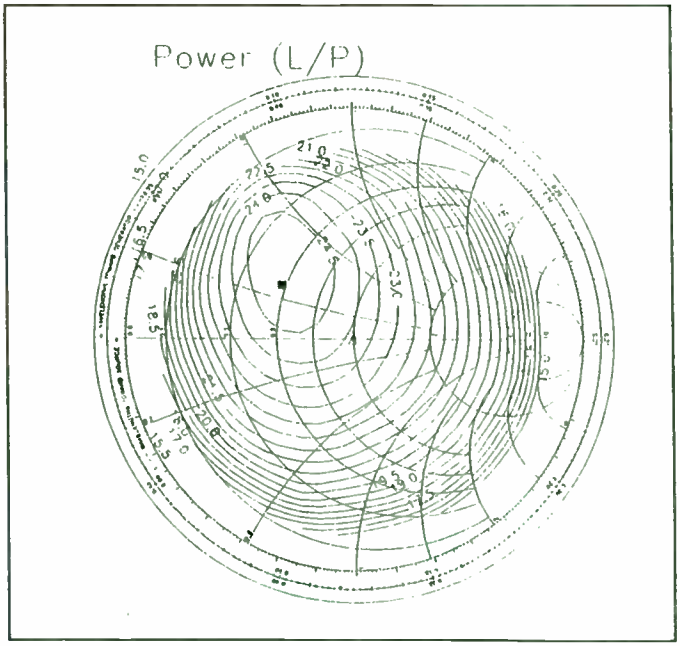


Figure 5. Load Pull in 50 Ω system, $\Gamma_{opt} \approx 0.56$

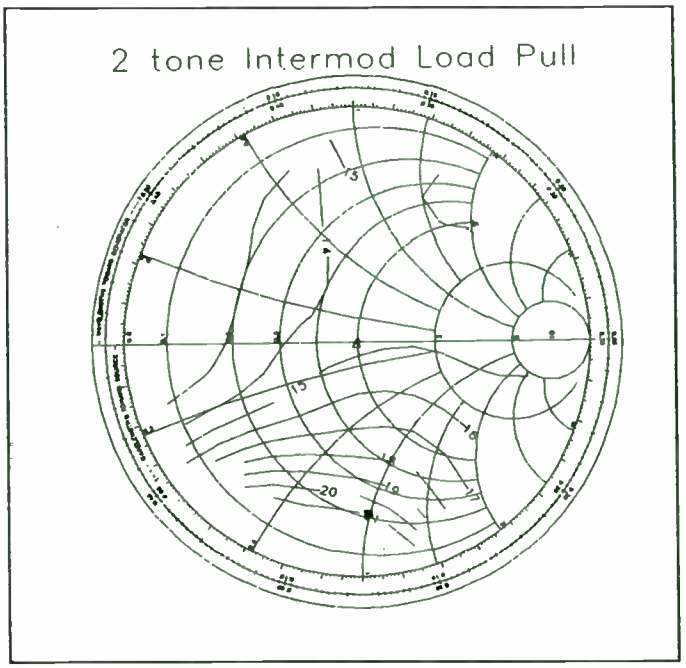


Figure 6. Intermod Load Pull

JDC spectral performance and higher order Intermod data can also be processed to ISO contours.
Using these plots the Design Engineer can synthesize a matching network that will generate the required performance. Overlapping different graphs, measured under the same conditions will also permit to make the best compromise in design between conflicting requirements.

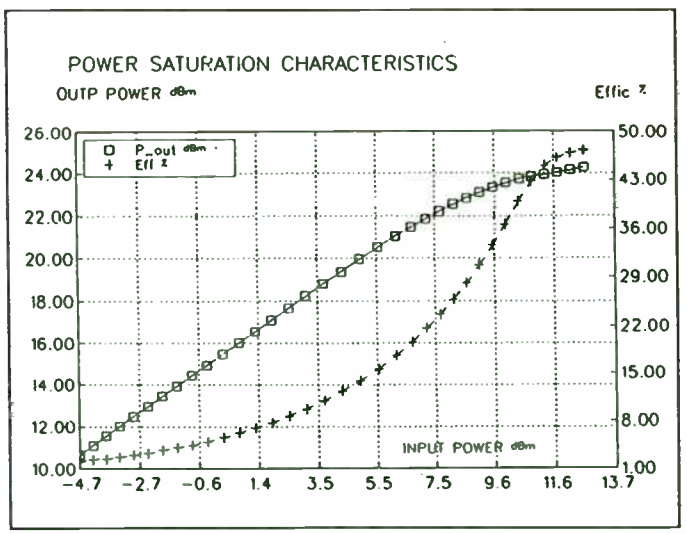


Figure 7. Power Transfer (Saturation)

JDC or Adjacent-Channel Leakage Load Pull
This test method has been developed in conjunction with our customers in the Digital Cellular Telecommunications business. It permits to optimize the performance of transceivers by measuring the ratio of the carrier-wave power integral to the integral of the power leaked within the upper (or lower) adjacent-channel bandwidth (figure 8).

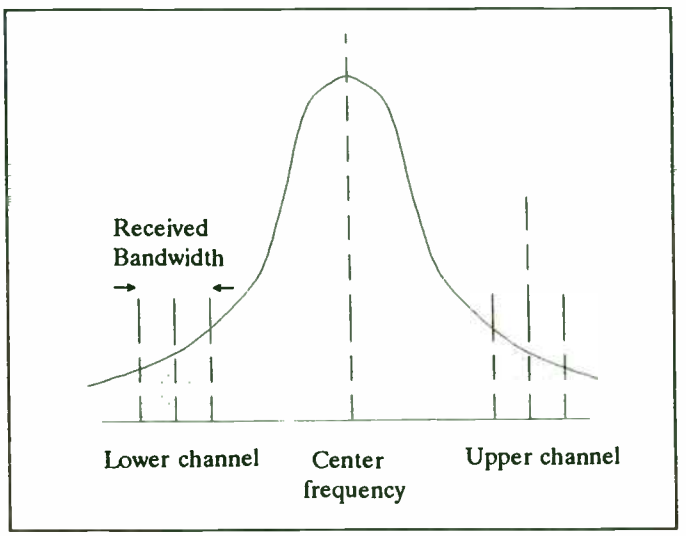


Figure 8. Adjacent channel leakage measurement

The MTS includes two types of measurement methods for this test:

- **Auto**: The MTS uses 'built in' software routines in the spectrum analyzers themselves, like the Advantest R3271 or Anritsu MS2602A and processes the final results as delivered by the analyzer. The only parameter setable by the MTS software in this case is the center frequency. The analyzer measures automatically the adjacent-channel power at preset conditions, in general at 50 and 100 kHz below and above the carrier.
- **Custom**: The MTS uses any type of spectrum analyzer, even those who do not support the JDC test method, and a MTS custom software that permits to set markers, sample the channel power at distinct windows and integrate signal power in order to generate equivalent results. In this case the User has control over the following parameters of the measurement procedure

JDC Test Parameter	Default Value
- Center frequency	Tuner Frequency
- Sideband 1 Offset	50 kHz
- Sideband 2 Offset	100 kHz
- Frequency Step between Samples	1 kHz
- Number of Samples (per sideband)	5
- Averaging Factor	2
- Settling Sweeps (before sample)	1

Table 3. Measurement Parameter settings for JDC Load Pull Test

Figure 9 shows an example of JDC contour tests made using the MTS custom test software and an HP-8562 spectrum analyzer.

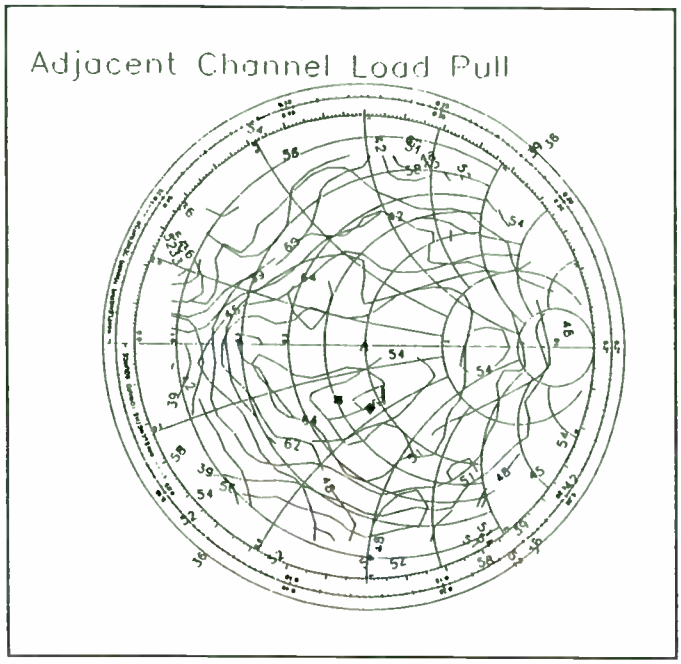


Figure 9. ISO contours of Adjacent channel leakage at 50 and 100 kHz sidebands

RF Power Amplifier Design

The MTS provides (as an option) a method to generate ISO power or Efficiency contour data compatible with Focus Microwave's network simulator **rf-PADS**, which can rapidly and accurately optimize high power and wideband amplifier stages, using power contour data only.

For this approach the User does not need to know more about the transistor than the external measurement conditions and have access to a set of small signal S-parameters, which are used to estimate only output-to-input feedback. All other optimization values, such as large signal input impedance and output power are generated directly by **numerical interpolation of power data**, measured using the MTS hard and software at rated power level. Figure 10 shows the principle of operation of rf-PADS.

The impedance conditions at the input are very important and they are an integral part of the data. This is also one of the main difficulties encountered by Engineers who try to design (approximately) power amplifiers using S-parameters.

In order to execute **rf-PADS**, only requires:

- the external program RF-PADS.EXE
- the Data Conversion program CNTDAT.EXE which generates ISO-power or ISO-efficiency files from premeasured Loaf Pull data files.

No other hardware or software components are required to use this option of MTS.

Rf-PADS includes lossy microstrip transmission line models, together with basic circuit elements such as capacitors, inductors and resistors. For the frequencies covered by MTS these elements are sufficient to design single stage power amplifiers.

Rf-PADS uses familiar .CKT type nodal network description as most other simulators. Due to the direct processing of measured data and second order interpolation techniques rf-PADS delivers **excellent accuracies** for even **saturated amplifier** stages.

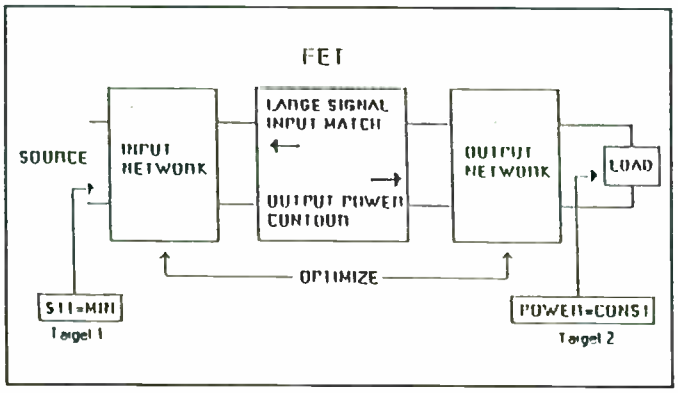


Figure 10. Principle of operation of rf-PADS

Power Amplifier Design procedure:

Rf-PADS uses a window type interface including a file editor.

- Step 1: Measure Load Pull data in a frequency range at prematched input conditions.
- Step 2: Convert the load pull data to rf-PADS contour files.
- Step 3: Generate a nodal description of a possible input and output matching network and set target performance.
- Step 4: Load the contour data into the circuit file and optimize the network parameters.

Due to direct data processing matching networks for constant output power and optimum input reflection can in general be found within minutes, using a normal -386 or -486 PC.

Figures 11-12 and table 4 show some results of high power amplifier designs together with the obtained design accuracy.

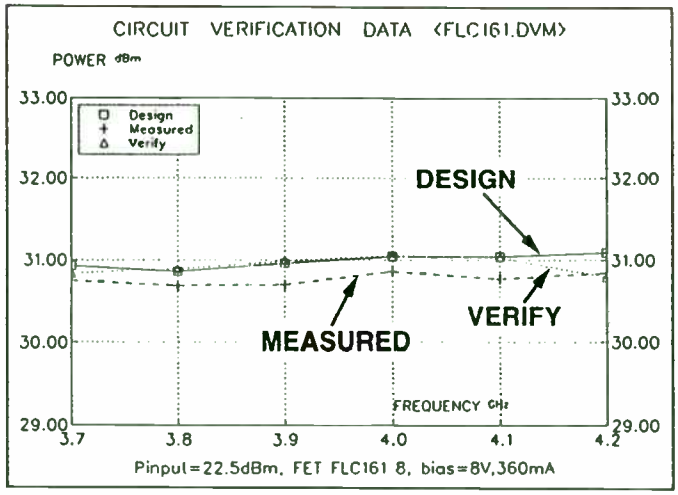


Figure 11. Comparison: Design - Measurement, Courtesy of CNET

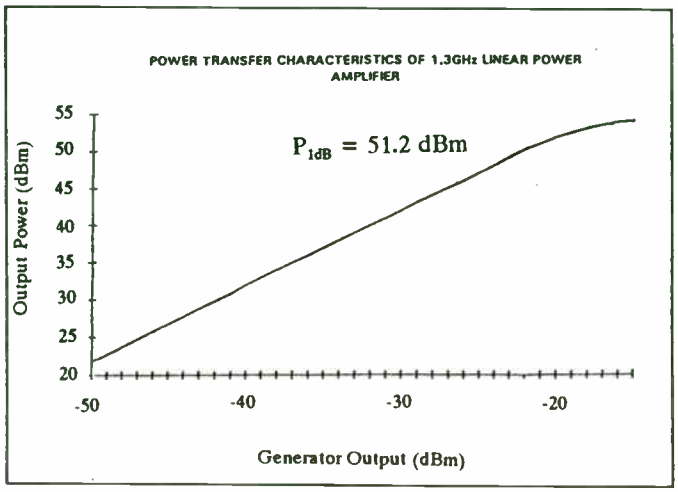


Figure 12. Pout(Pin) of High Power amplifier, Courtesy of ITS Electronics

Performance of 3.7 - 4.2 GHz Power Amplifier
FET FLC161WF (Fujitsu) Bias 8V, 360 mA
Input power 22.5dBm

GHz	Output Power dBm		Deviation dB	Gain dB	
	Design	Measure		Design	Measure
3.7	30.93	30.83	**0.10**	8.43	8.33
3.8	30.86	30.83	**0.03**	8.36	8.33
3.9	30.96	30.81	**0.15**	8.46	8.31
4.0	31.04	30.84	**0.20**	8.54	8.34
4.1	31.04	30.87	**0.17**	8.54	8.37
4.2	31.09	30.94	**0.15**	8.59	8.44

Table 4. Comparison: Design - Measurement

Conclusion

The Microwave Tuner System (MTS) of Focus Microwaves has been designed to respond to the key requirements of Test and Design Engineers in the PCN (Cellular Telecommunications) sector around 800 MHz to 3 GHz (optional to 4.2GHz).
These cover

- Accuracy (≈50 dB)
- Low Cost (1/3 of other systems)
- Versatility (Power - Intermod - Noise; auto and manual)
- Design Capability (0.2 dB accurate using power contours)

all integrated in an easy to calibrate and operate measurement and design workstation that can be setup around an IBM-PC and most popular GPIB instruments.

Acknowledgements

Many of the results and the measuring techniques used in the MTS software have been suggested or provided by our customers. We very much appreciate their support and contribution.

Noise Figure and Gain Measurement on High Speed Bipolar Junction Transistors

Wayne Jung
Tektronix Component Corporation, P.O. Box 500, M/S 59-355, Beaverton, OR 97077

Paul Van Halen
Portland State University, P.O. Box 751, EE Department, Portland, OR 97207

INTRODUCTION

It is a fact that no electronic system is completely free of random noise. Small voltage fluctuations due to noise are always occurring in electronic circuits because electrons are discrete and are constantly moving in time. The term "noise" originated with the study of high-gain audio-frequency amplifiers. When a fluctuating voltage or current generated in a device is amplified by an audio-frequency amplifier and the amplified signal is fed into a loudspeaker, the loudspeaker produces a hissing sound, hence the name noise. This descriptive term, noise, now refers to any spontaneous fluctuation in a system, regardless of whether an audible sound is produced.

Noise obscures low level electrical signals, therefore it can be a limiting factor on component and system performance. In a spectrum analyzer, for example, noise limits the sensitivity of the instrument--that is, the lowest-amplitude signal that the analyzer can detect and display. In a radar system, noise can obscure returns from a target, and limit the effective range of the radar. In digital communications, excessive noise can cause a high Bit-Error-Rate, resulting in the transmission and reception of false information.

This analysis studies the noise and gain characteristics of the basic component in an amplifier block, the transistor, in the microwave frequency range. After identifying the noise sources in the transistor, the effect of each noise source on the performance of the device is discussed. This study will provide spot noise figure and gain data for two high speed silicon bipolar transistors manufactured on two processes from Tektronix Microelectronic (SHPi and GST-1) . SHPi is Tektronix' latest generation Super-High frequency bipolar processes. The f_T of the transistor of interest (N16) of this process is measured to be close to $9GHz$ under the condition: $I_C = 9mA$ and $V_{CE} = 4V$. GST-1 (Giga-Speed Si-Bipolar Technology) is a high speed self-aligned double-polysilicon process. GST-1 is designed for the purpose of building high density, high performance circuits. The f_T for the G14V102, an N16 equivalent in the GST-1 process, is in the proximity of $12GHz$ with $I_C = 25mA$ and $V_{CE} = 4V$.

BASIC CONCEPTS

The measurement techniques used at microwave frequencies are different compared with low frequency methods. At lower frequencies, the properties of a circuit or system are determined by measuring voltages and currents. This approach is not applicable to microwave circuits since oftentimes these quantities are not uniquely defined. As a result, most microwave experimentation involves the accurate measurement of impedance and power rather than voltage and current.

S-PARAMETERS

Introduction

Linear networks can be completely characterized by parameters measured at the network terminals without regard to the contents of the network. Once the parameters of a network have been determined, its behavior in any external environment can be predicted, again without regard to the specific contents of the network.

A two-port device can be described by a number of parameter sets. Hybrid, admittance, and impedance parameters sets are often used at low frequency analysis. Moving to higher frequencies, some problems arise:

1. Equipment is not readily available to measure total voltage and total current at the ports of the network. The voltage measured will not be the same as the voltage at the ports of the network if the length of the transmission line is comparable with the wavelength of the test signal.
2. Active devices, such as transistors, very often will not be short or open circuit stable.
3. Parasitics in active devices may cause unwanted oscillations.

Some other sets of parameters are necessary to overcome these problems. Hence, dissipating (resistive) loads are used in measurements to minimize parasitic oscillations. In addition, the concept of traveling waves rather than total voltages and currents is used to describe the networks. Voltage, current, and

power can be considered to be in the form of waves traveling in both directions along a transmission line (Fig. 1). A portion of the waves incident on the load will be reflected if the load impedance Z_l is not equal to the characteristic impedance of the transmission line Z_o. This is analogous to the concept of maximum power transfer. A source will deliver maximum power to a load (no reflection from the load) if the load impedance is equal to the source impedance.

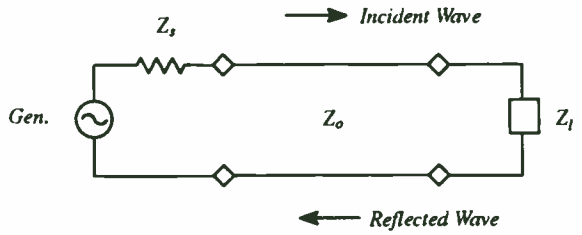


Figure 1. Traveling waves.

Reflection Coefficients

The reflection coefficient Eq. (1) is a mathematical representation of the reflected voltage wave with respect to the incident voltage wave at a specified port of a circuit.

$$\Gamma = \frac{Reflected\ Wave}{Incident\ Wave} \tag{1}$$

Eq. (2) defines the load reflection coefficient in terms of the load impedance referenced to the characteristic impedance of the transmission line.

$$\Gamma_l = \frac{Z_l - Z_o}{Z_l + Z_o} \tag{2}$$

And similarly the source reflection coefficient is

$$\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o} \tag{3}$$

It is well known that maximum power transfer occurs when the impedance of the source matches the impedance of the load. For $\Gamma = 0$, maximum power transfer occurs, and Γ equals unity when there is no power transfer since then all the incident power is reflected back.

S-parameters

For a two-port network shown in Fig. 2, the following new variables are defined [1]:

$$a_1 = \frac{E_{i1}}{\sqrt{Z_o}} \qquad a_2 = \frac{E_{i2}}{\sqrt{Z_o}}$$
$$b_1 = \frac{E_{r1}}{\sqrt{Z_o}} \qquad b_2 = \frac{E_{r2}}{\sqrt{Z_o}}$$

where E_i and E_r are the incident voltage wave and reflected voltage wave respectively. Notice that the square of the magnitude of these new variables has the dimension of power.

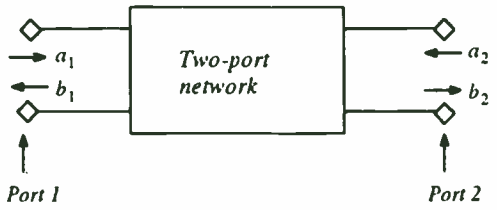


Figure 2. Incident and reflected waves in a two-port network.

By measuring incident and reflected power from a two-port network, we avoid the primary problem in microwave measurement, which is the voltage variation along the transmission line between the device under test and the test equipment. A new set of parameters relate these four waves in the following fashion:

$$b_1 = S_{11}a_1 + S_{12}a_2$$
$$b_2 = S_{21}a_1 + S_{22}a_2$$

where

$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}$ input reflection coefficient with the output matched

$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}$ forward transmission coefficient with the output matched

$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}$ output reflection coefficient with the input matched

$S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}$ reverse transmission coefficient with the input matched

This set of new parameters is called "scattering parameters," since they relate those waves scattered or reflected from the network to those waves incident upon the network. These scattering parameters are commonly referred to as s-parameters.

Although s-parameters completely characterize a linear network, sources and loads need to be attached before the network can be used. The new input reflection coefficient Γ_{in} an the new output reflection coefficient Γ_{out} take the form [2]:

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_l}{1 - S_{22}\Gamma_l} \tag{4}$$

and

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \tag{5}$$

The second terms in the above equations show the interaction between the two ports. If S_{12} is equal to zero, the two-port network is called unilateral because whatever happens to one port does not affect the other port.

For a bilateral linear two-port network, the input reflection coefficient is not just function of S_{11}. It is a function of Γ_l as well. Therefore, the effect of the load must be considered when

an input match circuit is built to match the signal source. Also, Γ_{out} is a function of the s-parameters and Γ_s, and this should be kept in mind when an output matching network is built to have all the power delivered to the load.

Smith Chart

The Smith chart (Fig. 3) is a design and analysis tool that can provide insight into the impedance and reflection characteristics of a circuit. The chart is a graphical representation that provides a transformation between the complex reflection coefficient in a polar format to the real and imaginary parts of the impedance, Eq. (2) and Eq. (3). The chart has the property of being able to graphically display the entire range of real and imaginary values of input impedances of a network. Therefore, on a Smith chart a point simultaneously represents three different things. Depending on the coordinate system used as a reference, they are: reflection coefficient, impedance, and admittance. Impedance matching circuits can be easily and quickly designed using the Smith chart.

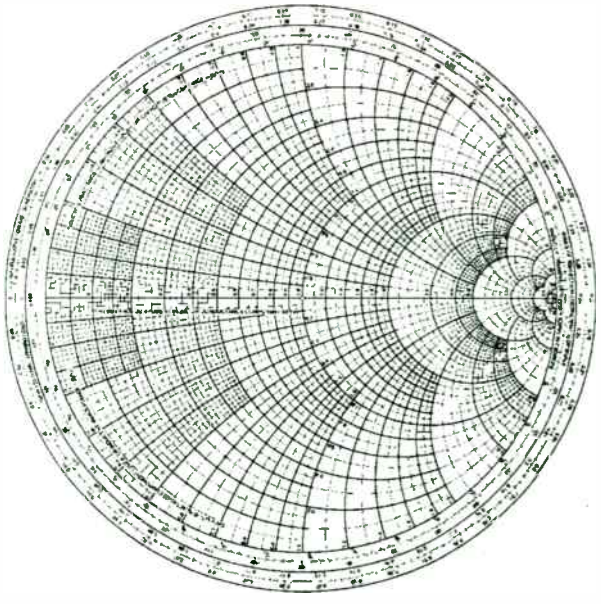


Figure 3. The Smith chart.

NOISE FIGURE

To compare the performance of high frequency devices, an important figure of merit is the noise factor/figure, which is a measure of how the thermal noise and the shot noise generated in semiconductors degrade the signal-to-noise ratio. The noise factor F of the network is defined as the ratio of the available signal-to-noise ratio at the signal generator terminals to the available signal-to-noise ratio at its output terminals.

$$F = \frac{S_i/N_i}{S_o/N_o} \qquad (6)$$

The current definition for noise factor applies whether it is linear or logarithmic. The term "noise factor (F)" is used when referring to the ratio in linear terms. The term "noise figure (NF)" is used when referring to the ratio in logarithmic terms.

$$NF = 10\ log_{10}F \qquad (7)$$

From the definition in Eq. (6) the ideal noise factor is unity or noise figure is $0dB$, where there is no degradation in signal-to-noise ratio after the signal passes through the network.

Noise figure is a parameter that applies both to components and systems. The overall system performance can be predicted from the noise figure of the components that go into it. For a number of networks in cascade, as shown in Fig. 4, the system spot noise factor is given in terms of the component spot noise factors (F_i) and available gain (G_i) by

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots \qquad (8)$$

From Eq. (8), the significance of the first stage gain and noise factor are evident. The first stage in the chain has the most significant contribution to the total noise factor of the chain. The total noise factor of the system is at least equal to the noise factor of the first stage. If the first stage gain is significantly large, then the noise contributions from the succeeding stages will be small.

Figure 4. Noise figure of network in cascade.

AVAILABLE POWER GAIN

A signal generator with an internal impedance R_s ohms and output voltage V_s volts can deliver $\frac{V_s^2 R_l}{(R_s + R_l)^2}$ watts into a resistance of R_l ohms. This power is maximum and equal to $\frac{V_s^2}{4R_s}$ when the output circuit is matched to the generator impedance, that is when $R_l = R_s$. Therefore, $\frac{V_s^2}{4R_s}$ is called the available power of the generator, and it is, by definition, independent of the impedance of the circuit to which it is connected. The output power is smaller than the available power when R_l is not equal to R_s, since there is a mismatch loss. In amplifier input circuits a mismatch condition may be beneficial due to the fact that it may decrease the output noise more than the output signal. It is the presence of such mismatch conditions in amplifier input circuits that makes it desirable to use the term available power.

The symbol S_g will be used for the available signal power at the output terminals of the signal generator shown in Fig. 5. The

symbol S will be used for the available signal power at the output terminals of the two-port network.

The gain of the network is defined as the ratio of the available signal power at the output terminals of the network to the available signal power at the output terminals of the signal generator. Hence

$$G_A = \frac{\textit{power available from the network}}{\textit{power available from the source}} = \frac{S}{S_g} \quad (9)$$

Note that while the gain is independent of the impedance which the output circuit presents to the network, it does depend on the impedance of the signal generator.

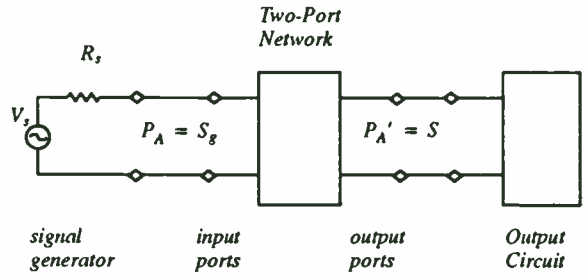


Figure 5. Available power gain of a two-port network.

NOISE FIGURE MEASUREMENT

NOISE IN BIPOLAR TRANSISTOR

The noise sources in a bipolar junction transistor are categorized into four major types:

1. Flicker noise;
2. Burst noise;
3. Shot noise;
4. Thermal noise.

Flicker Noise

Flicker noise is caused by traps associated with contamination and crystal defects in the emitter-base depletion layer. It is associated with a flow of direct current and displays a spectral density of the form [3]:

$$\overline{i^2} = \frac{K_F I^{A_F}}{f} \Delta f \quad (10)$$

where K_F is a constant for a particular device, A_F is a constant between 0.5 and 2, and Δf is the bandwidth in Hertz. This expression shows that the noise spectral density has a ($1/f$) frequency dependence, therefore, it is also called $1/f$ noise. Since flicker noise is most significant at low frequencies, it is not discussed here.

Burst Noise

It has been found experimentally that the low frequency noise spectrums of some bipolar transistors show a different frequency dependence than flicker noise. This could be the result of the existence of burst noise. It is caused by the imperfection in the crystal structure. The power spectrum of such signal is given by [4]:

$$\overline{i^2} = \frac{K_B I}{1 + \left(\frac{\pi f}{2k}\right)^2} \Delta f \quad (11)$$

where K_B is a technological dependent constant for a particular device, and k is the mean repetition rate of the signal. This noise is insignificant at microwave frequencies because it is inversely proportional to f^2, and it will not be addressed further.

Shot Noise

Shot noise is due to generation and recombination in the pn junction and injection of carriers across the potential barriers, therefore it is present in all semiconductor diodes and bipolar transistors. Each carrier crosses the junction in a purely random fashion. Thus the current I, which appears to be a steady current, is, in fact, composed of a large number of random independent current pulses. The fluctuation in I is termed shot noise and is generally specified in terms of its mean-square variation about the average value I_D, and it is represented by [3]:

$$\overline{i^2} = 2qI_D\Delta f \quad (12)$$

where q is the electronic charge ($1.6 \times 10^{-19}C$). Eq. (12) shows that the noise spectral density is independent of frequency. In a transistor, there are two such noise sources. They are the shot noise in the emitter-base junction (i_b) and in the collector-base junction (i_c).

Thermal Noise

Thermal noise is due to the random thermal motion of electrons in a resistor, and it is unaffected by the presence or absence of direct current, since typical electron drift velocities in a conductor are much less than electron thermal velocities. As the name indicates, thermal noise is related to absolute temperature T.

In a resistor R, thermal noise can be represented by a series voltage generator $\overline{v^2}$. It is represented by [3]:

$$\overline{v^2} = 4kTR\Delta f \quad (13)$$

where k is Boltzmann's constant ($1.38 \times 10^{-23}J/K$), and T is temperature in Kelvin. Like shot noise, thermal noise is also independent of frequency. Thermal noise is a fundamental physical phenomenon and is present in any linear passive resistor.

Equivalent Circuit

Fig. 6 is the full small-signal equivalent circuit, including noise sources for the bipolar transistor at high frequency [3]. Three noise sources are evident from the figure. They are thermal noise from the series input resistance (r_b) and shot noise due to the base and collector currents (I_b, I_c). and their values are:

$$\overline{v_b^2} = 4kTr_b\Delta f \tag{14}$$

$$\overline{i_b^2} = 2qI_b\Delta f \tag{15}$$

$$\overline{i_c^2} = 2qI_c\Delta f \tag{16}$$

The resistors r_π and r_o in Fig. 6 are equivalent circuit elements, not physical resistors, and they do not produce any thermal noise.

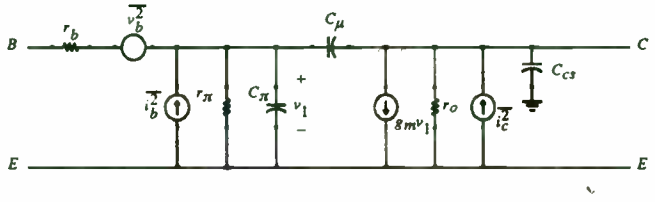


Figure 6. Bipolar transistor small-signal equivalent circuit with noise sources.

Neither thermal noise nor shot noise is frequency dependent, and both exhibit uniform noise output through the entire useful frequency range of the transistor. The internal gain of the transistor does vary with frequency, however, and it falls off as frequency increases. As a result the noise figure begins to rise when the reduction in gain becomes appreciable. Since the power gain falls inversely as frequency squared, the noise figure rises as frequency squared, or 6*dB* per octave [5]. Fig. 7 graphically shows the noise figure of a N16 transistor with $V_{BE} = 0.8V$, $V_{CE} = 4V$, and $R_s = 50\Omega$ in common emitter configuration.

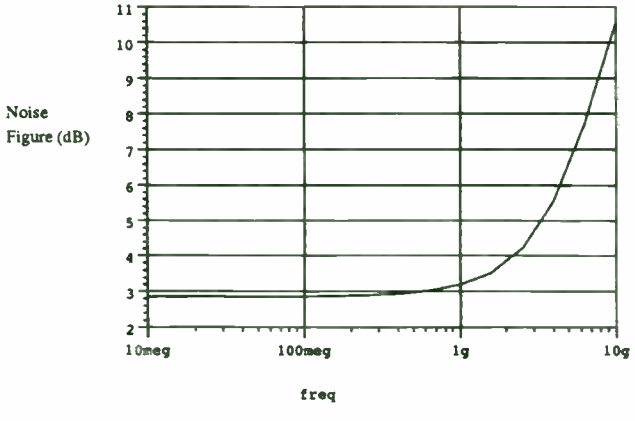


Figure 7. Noise figure vs. frequency for N16.

DETERMINATION OF NOISE PARAMETERS

Definition

As defined in 1960 by the Institute of Radio Engineers Subcommittee on Noise [6], the noise figure depends upon the internal structure of the transducer and upon its input termination,. but not upon its output termination. Thus, the noise performance of a transducer is meaningfully characterized by its noise figure only if the input termination is specified. The noise factor F of any linear transducer, at a given operating point and input frequency, varies with the admittance Y_s of its input termination in the following manner [7]:

$$F = F_o + \frac{R_n}{G_s} | Y_s - Y_o |^2 \tag{17}$$

where G_s is the real part of Y_s, and the parameters F_o, Y_o, and R_n characterize the noise properties of the transducer and are independent of its input termination. Thus the noise performance of a transducer can be meaningfully characterized for all input terminations through specification of the parameters F_o, Y_o, and R_n.

The "optimum noise factor" F_o, at the given transducer operating point and frequency, is the lowest noise factor that can be obtained through adjustment of the source admittance Y_s. The "optimum source admittance" Y_o is that particular value of source admittance Y_s which results in optimum noise factor F_o. The parameter R_n is positive and has the dimensions of a resistance. It is called the equivalent noise resistance. This parameter appears as the coefficient of the $| Y_s - Y_o |^2$ term in the general expression for F and, therefore, characterizes the rapidity with which F increases above F_o as Y_s departs from Y_o.

The parameters F_o, Y_o, and R_n can be calculated if the noise theory of the transducer is known or, alternatively, can be determine empirically from noise measurements.

Two methods of computer-aided determination of noise parameters have been reported in the literature.

One of them, Kokyczka et al [8], can be thought of as an automatic version of the graphic procedure suggested by the Institute of Radio Engineers [6], which required tedious and time-consuming adjustment of some input termination admittances with constant real part and of some other with constant imaginary part.

The other one, Lane [9], is an application of the least-squares method, which reduces the determination of noise parameters to the solution of a four linear equation system, obtained as fit of noise figures measured for different source admittances.

Noise Parameters Computation

Ten sets of data for the N16 and the G14V102 are obtained with different source and are tabulated in table I and will be used to calculate the noise parameters of the devices.

TABLE I

NOISE FIGURE MEASURED AT VARIOUS SOURCE IMPEDANCES
WITH V_{BE}=0.8V, V_{CE}=4V AT F=1GHZ

G_s mmhos	B_s mmhos	NF dB	
		N16	G14V102
6.33	-2.26	2.258	2.724
17.8	6.24	3.316	2.675
24.9	6.64	3.783	2.830
18.4	24.1	5.514	4.192
7.14	-7.14	2.702	2.697
13.5	-18.9	4.446	3.205
44.9	-28.1	5.751	3.948
14.2	-1.02	2.714	2.365
15.7	-3.92	2.861	2.362
22.8	-7.12	3.509	2.585

After the noise figure data for different source impedances are obtained, the minimum noise figure, the optimum source resistance, and the equivalent noise resistance can be determined. Eq. (17) is rewritten to eliminate the magnitude sign:

$$F = F_o + \frac{R_n}{G_s} [(G_s - G_o) + j(B_s - B_o)]^2 \quad (18)$$

In order to use a readily available subroutine for solving simultaneous equation solution, Eq. (18) is transformed to a form that is linear with respect to four new parameters A, B, C, and D [11].

$$F = A + BG_s + \frac{C + BB_s^2 + DB_s}{G_s} \quad (19)$$

where

$$F_o = A + \sqrt{4BC - D^2} \quad (20)$$

$$R_n = B \quad (21)$$

$$G_o = \frac{\sqrt{4BC - D^2}}{2B} \quad (22)$$

$$B_o = \frac{-D}{2B}. \quad (23)$$

In principle, four measurements of noise factor from different source admittances will determine the four real numbers (F_o, R_n, G_o, and B_o). Eq. (18) becomes overdetermined if more than four measurements are taken, but by minimizing the square of error as expressed in Eq. (24), more than four measurements can be used to find those parameters which give the best least squares fit to Eq. (18). It has been shown that only slight variations of noise parameters occur versus redundancy if the number of data sets processed is greater than 7 [10]. A least-squares fit of the ten sets of noise figure data noise from table I to Eq. (19) is sought: therefore, the following error criterion is established [9]:

$$\varepsilon = \frac{1}{2} \sum_{i=1}^{n} W_i \left[A + B \left(G_i + \frac{B_i^2}{G_i} \right) + \frac{C}{G_i} + \frac{DB_i}{G_i} - F_i \right]^2 \quad (24)$$

where W_i is the weighting factor to be used if certain data are known to be less reliable than the average.

Base on the noise figure data in table I, the noise parameters for the N16 and G14V102 transistors are obtained.

TABLE II

NOISE PARAMETERS OF N16 AND G14V102 TRANSISTORS
WITH V_{BE}=0.8V, V_{CE}=4V AT F=1GHZ

	Fo dB	Rn ohms	Go mmhos	Bo mmhos
N16	2.25	42.67	6.31	-1.66
G14V102	2.34	21.32	13.12	-3.43

NOISE FIGURE CIRCLES

To plot noise figures on a Smith chart, Y_s and Y_o are expressed in terms of reflection coefficients Γ_s and Γ_o, and Eq. (17) becomes:

$$F = F_o + \frac{4R_n |\Gamma_s - \Gamma_o|^2}{(1 - |\Gamma_s|^2)|1 + \Gamma_o|^2} \quad (25)$$

This equation can be used to seek Γ_s for a given noise figure. To determine a family of noise figure circles, an intermediate noise figure parameter, N_i is defined [1].

$$N_i = \frac{|\Gamma_s - \Gamma_o|^2}{1 - |\Gamma_s|^2} \quad (26)$$

Eq. (26) is then transformed to:

$$\left| \Gamma_s - \frac{\Gamma_o}{1 + N_i} \right|^2 = \frac{N_i^2 + N_i (1 - |\Gamma_o|^2)}{(1 + N_i)^2} \quad (27)$$

Eq. (27) is recognized as a family of circles with N_i as a parameter. The center and radius of the circle can be found from:

$$C_{F_i} = \frac{\Gamma_o}{1 + N_i} \quad (28)$$

and

$$R_{F_i} = \frac{1}{1 + N_i} \sqrt{N_i^2 + N_i (1 - |\Gamma_o|^2)} \quad (29)$$

Eq. (26), Eq. (28), and Eq. (29) show that when $F_i = F_o$, then

$N_i = 0$, $C_{F_o} = \Gamma_o$, and $R_{F_o} = 0$. That is, the center of the F_o circle is located at Γ_o with zero radius. From Eq. (28), the centers of the other noise figure circles are located along the Γ_o vector.

A set of constant noise figure circles for the N16 is shown in Fig. 8. This plot gives information about the noise figure of the device for different source impedances at $1GHz$. A $F_i dB$ noise figure circle on the plot specifies the values of source impedance at which the device will produce a noise figure of F_i. This set of curves show that the minimum noise figure, $F_o = 2.25dB$ is obtained when $\Gamma_s = \Gamma_o = 0.52\angle 10.6°$, and at point A, $\Gamma_s = 0.31\angle 80°$ produces $F_i = 3.25dB$. One point of interest is on the center of the Smith chart, which corresponds to a source impedance of 50Ω. The noise figure at this point is what we can have if the device is put into a 50Ω environment without any tuning circuitry at the input ports.

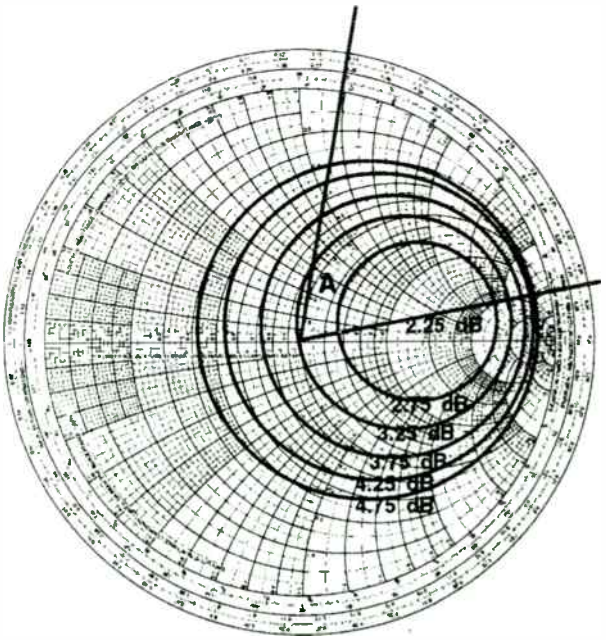


Figure 8. Constant noise figure circles for N16 with V_{BE} = 0.8V, V_{CE} = 4V at f=1GHz.

COMPARISON

From the data obtained, the equivalent noise resistance (R_n) of the G14V102 is smaller than the one of the N16. This is verified by wider spacing of noise figure circles of the G14V102 as compare to the ones of the N16 (Fig. 9 vs. Fig. 10). The lower of R_n will result in reduced sensitivity of the noise figure to changes in source impedance. Therefore, a circuit designer can have more freedom on choosing source impedances for better power gain and/or better input matching for a given noise figure.

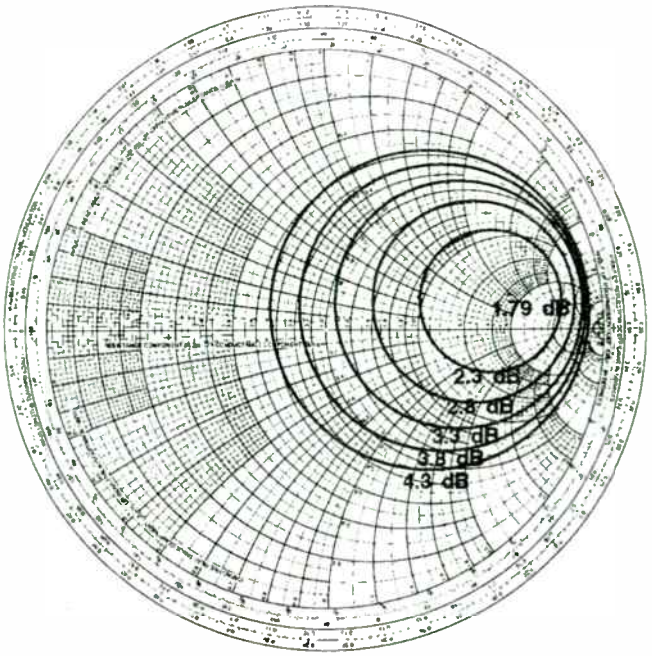


Figure 9. Constant noise figure circles for N16 with V_{BE}=0.76V, V_{CE}=4V at f=900MHz.

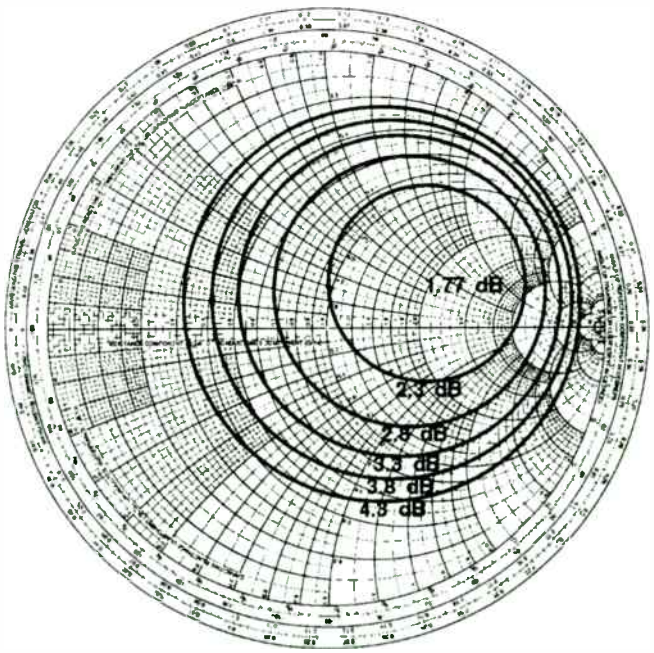


Figure 10. Constant noise figure circles for G14V102 with V_{BE}=0.76V, V_{CE}=4V at f=900MHz.

GAIN MEASUREMENT

GAIN AND STABILITY

Stability Criteria

As the operating frequency of the transistor is being pushed upward, the transistor is more prone to unwanted oscillation due to parasitic elements. The necessary conditions for stability of a two-port device like bipolar transistors, had been studied by Kurokawa[12], Bodway[13], and Woods[14]. In terms of s-parameters, they are:

$$| \Delta | = | S_{11}S_{22} - S_{12}S_{21} | < 1 \tag{30}$$

and

$$K = \frac{1 - | S_{11} |^2 - | S_{22} |^2 + | \Delta |^2}{2 | S_{12}S_{21} |} > 1 \tag{31}$$

Two-port devices that meets the above criteria are unconditionally stable for any positive source and load impedance.

Stability Circles

To maximize the gain, we must conjugately match the input and the output. For unconditionally stable two-ports networks there is no unwanted oscillation to worry about. But for those networks which cannot meet the above stability criteria, we will have to look at what might happen to the network in terms of stability--will the amplifier oscillate with certain values of impedance used in the matching process?

In a two-port network, oscillations are possible when either the input or output port presents a negative resistance, since noise generated in the adjoining network enters the port, the negative resistance generates more noise rather than dissipating the incident noise, and some of this generated noise combines with the incoming noise to input more noise. Negative resistances correspond to the points outside the Smith chart, which imply either $| \Gamma_{in} | > 1$ or $| \Gamma_{out} | > 1$. Therefore, we have the boundary for the input and output stability circles defined:

$$| \Gamma_{in} | = \left| S_{11} + \frac{S_{12}S_{21}\Gamma_l}{1 - S_{22}\Gamma_l} \right| = 1 \tag{32}$$

and

$$| \Gamma_{out} | = \left| S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \right| = 1 \tag{33}$$

Solving for the values of Γ_l and Γ_s in Eq. (32) and Eq. (33) shows that the solutions for Γ_l and Γ_s lie on circles [2]. The radius and center of the input stability circles are:

$$r_s = \left| \frac{S_{12}S_{21}}{| S_{11} |^2 - | \Delta |^2} \right| \tag{34}$$

$$C_s = \frac{(S_{11} - \Delta S_{22}^*)^*}{| S_{11} |^2 - | \Delta |^2} \tag{35}$$

and the radius and center of the output stability circles are:

$$r_l = \left| \frac{S_{12}S_{21}}{| S_{22} |^2 - | \Delta |^2} \right| \tag{36}$$

$$C_l = \frac{(S_{22} - \Delta S_{11}^*)^*}{| S_{22} |^2 - | \Delta |^2} \tag{37}$$

where

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{38}$$

Having measured the s-parameters of a two-port device at one frequency, Eq. (34) to Eq. (37) can be evaluated, and plotted on a Smith chart. Fig. 11 illustrates the graphical construction of the stability circles where $| \Gamma_{in} | = 1$. On one side of the stability circles boundary, in the Γ_l plane, we will have $| \Gamma_{in} | < 1$ and on the other side $| \Gamma_{in} | > 1$.

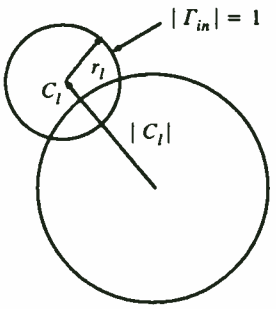


Figure 11. Stability circles construction on a Smith chart.

To determine which area represents the stable operating condition, we make $\Gamma_l = 0$, which is to terminate the output port with a 50Ω load through a 50Ω transmission line. This represents the point at the center of the Smith chart. Under these conditions,

$$| \Gamma_{in} | = | S_{11} | \tag{39}$$

For N16, the magnitude of S_{11} measured is less than unity, therefore, the center of the Smith chart represents a stable operating point. That is, the shade area on Fig. 12 represents $| \Gamma_{in} | < 1$. The same procedure applies for finding the output stability region.

When the input and output stability circles lie completely outside the Smith chart the network is called unconditionally stable for all Γ_s and Γ_l. This comes from the fact that no matter what positive termination is put at the input or output of the network $| \Gamma_{in} |$ and $| \Gamma_{out} |$ will be always less than unity.

Gain Circles

S-parameters can be used to predict the available power gain of a transistor for any input termination Γ_s. This available gain is

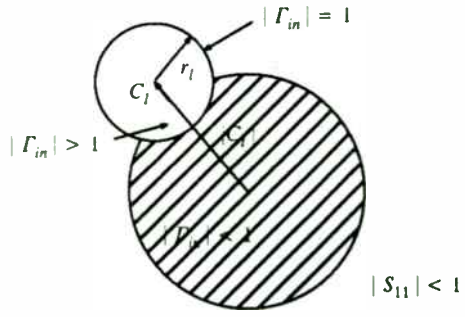


Figure 12. Stability region for Γ_s.

that gain achieved when a transistor is driven from some source reflection Γ_s while terminated with a load impedance equal to Γ_{out}^* (matched output). The available power gain in terms of reflection coefficients is [1]:

$$G_A = \frac{(1 - |\Gamma_s|^2)}{|1 - S_{11}\Gamma_s|^2} |S_{21}|^2 \frac{1}{(1 - |\Gamma_{out}|^2)} \tag{40}$$

where

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \tag{41}$$

Because G_A is a function of the source reflection coefficient, constant available power gain circles can be plotted on a Smith chart together with the constant noise figure circles, and the trade-off result between gain and noise figure can be analyzed.

For a given gain G_A, the radius R_a and the center C_a of the circle can be calculated using the relations [2]

$$g_a = \frac{G_A}{|S_{21}|^2} \tag{42}$$

$$C_1 = S_{11} - \Delta S_{22}^* \tag{43}$$

$$R_a = \frac{\sqrt{1 - 2K|S_{12}S_{21}|g_a + |S_{12}S_{21}|^2 g_a^2}}{|1 + g_a(|S_{11}|^2 - |\Delta|^2)|} \tag{44}$$

and

$$C_a = \frac{g_a C_1^*}{1 + g_a(|S_{11}|^2 - |\Delta|^2)} \tag{45}$$

For a given G_A, the constant available power gain circle can be plotted. All Γ_s on this circle produce the gain G_A.

COMPARISON

Gain circles, stability circles, and noise figure circles for the N16 and the G14V102 are plotted in Fig. 13 and Fig. 14. Under the conditions $V_{BE} = 0.76V$, $V_{CE} = 4V$ at $f = 900MHz$, these two transistors can provide similar power gains. Looking at these figures, one can see that a N16 will provide a gain of $16dB$ and a noise figure of $3.5dB$ with 50Ω source impedance, while a G14V102 will provide about the same gain but with a superior noise figure of $2.4dB$.

CONCLUSION

Noise is created by many physical processes which cannot be avoided. Living with noise means we must be able to measure and predict it. The noise sources in a bipolar transistor have been identified. At microwave frequencies they are thermal noise due to base resistance and shot noise from the base and collector currents. The base resistance consists of two parts. The external base resistance, R_{b_i}, is the resistance of the path between the base contact and the edge of the emitter diffusion. The active base resistance, R_{b_a}, is that resistance between the edge of the emitter and the site within the base region at which the current is actually flowing. R_{b_i} can be reduced by decreasing the separation between the base and the emitter. This method is straightforward, but it is technology-limited. While the effect of current crowding in the base at high current level will reduce the effect of R_{b_a}, but more shot noise will be generated by the higher current.

The technique of measuring noise parameters (F_o, R_n, G_o, and B_o) of a bipolar transistor has been presented. This same measurement technique can also be employed to measure the noise parameters of a general two-port network. Power gain information is obtained through s-parameter manipulation. The noise figure and available power gain data are plotted on Smith Chart to give a clear view of how a particular device will perform in various source impedances.

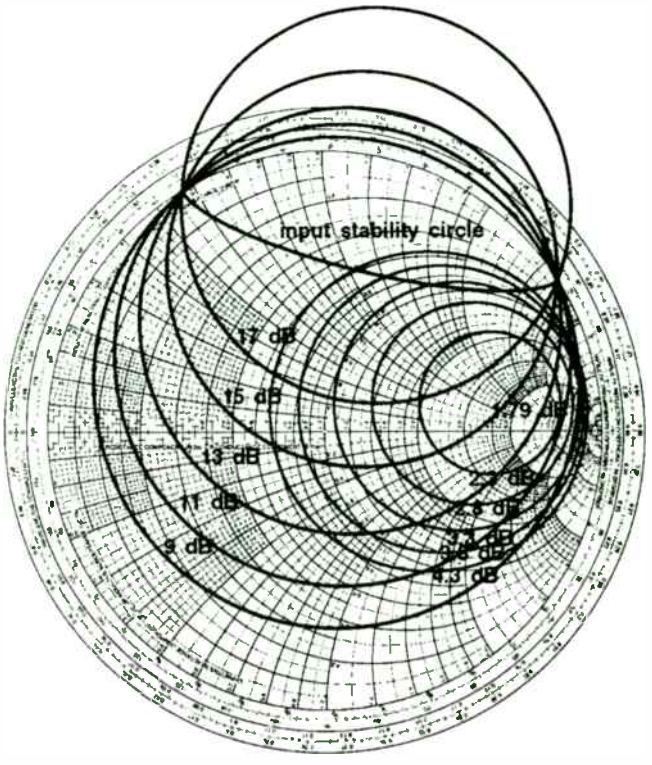


Figure 13. Gain and noise figure circles for N16 at f=900MHz.

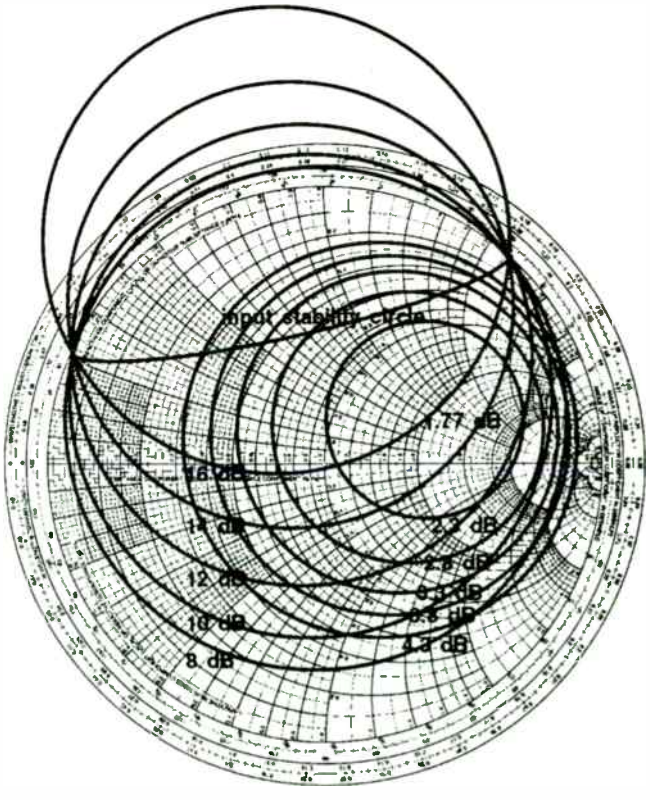


Figure 14. Gain and noise figure circles for G14V102 at f=900MHz.

REFERENCE

[1] "S-parameter Design," Hewlett-Packard Application Note 154, Mar. 1990.

[2] Guillermo Gonzalez, *Microwave Transistor Amplifiers*, Prentice-Hall, New Jersey, 1984.

[3] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 2nd ed., John Wiley & Sons, New York, 1984.

[4] Z. Y. Chang and M. C. Willy, *Low-Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies*, Kluwer Academic Publishers, Norwell, MA, 1991.

[5] Harry F. Cooke, "Transistor Noise Figure," *Solid State Design*, pp. 37–42, Feb. 1963.

[6] H. A. Haus *et al.*, "IRE Standards on Methods of Measuring Noise in Linear Twoports, 1959," *Proc. IRE*, vol. 48, pp. 60–68, Jan. 1960.

[7] H. A. Haus *et al.*, "Representation of Noise in Linear Twoports," *Proc. IRE*, vol. 48, pp. 69–74, Jan. 1960.

[8] W. Kokyczka, A. Leupp, and M. J. O. Strutt, "Computer-aided determination of two-port noise parameters (CADON)," *Proc. IEEE*, vol. 58, pp. 1850–1851, Nov. 1970.

[9] Richard Q. Lane, "The Determination of Device Noise Parameters," *Proc. IEEE*, vol. 57, pp.1461–1462, Aug. 1969.

[10] G. Garuso and M. Sannino, "Computer-Aided Determination of Microwave Two-port Noise Parameters." *IEEE Transactions on Microwave Theory and Techniques*, vol. 26, pp. 639–642, Sept. 1978.

[11] H. Fukui, "The Noise Performance of Microwave Transistors," *IEEE Transactions on Electron Devices*, vol. 13, pp. 329–341, Mar. 1966.

[12] K. Kurokawa, "Power Waves and the Scattering Matrix," *IEEE Transactions on Microwave Theory and Techniques*, vol. 13, pp. 194-202, Mar. 1965.

[13] G. E. Bodway, "Two Port Power Flow Analysis Using Generalized Scattering Parameters," *Microwave Journal*, May 1967.

[14] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of S Parameters," *IEEE Transactions on Circuits and Systems*, Feb. 1976.

[15] Ronald W. Kruse, "Microwave Design Using Standard Spice," *Microwave Journal*, pp. 164–171, Nov. 1988.

THE CURRENT-FEEDBACK OP AMP
A HIGH-SPEED BUILDING BLOCK

Anthony D. Wang
Burr-Brown Corporation
6730 South Tucson Boulevard
Tucson, AZ 85706-7087

Although current-feedback amplifiers (CFAs) have been in use for quite some time, there is a reluctance to view them in the same light as voltage-feedback amplifiers (VFAs). For instance, the gain-bandwidth curve of VFAs has a parallel in a transimpedance-bandwidth curve for CFAs. This parameter can be used to determine the closed-loop behavior of the CFA in the same way that GBW can for the VFA. Not all the fault is with the users - the amplifier manufacturers have not standardized the CFA characterization as they have done with VFAs. This paper describes the CFA and its behavior in an intuitive manner.

HISTORICAL PERSPECTIVE

The term "operational amplifier", or "op amp" in typical engineering shorthand, has generally been associated with the transistorized voltage-feedback amplifier. It is becoming more acceptable now to include the current-feedback amplifier in the same category.

Interestingly enough, the basic architecture for the CFA might have predated the VFA although it was not until the 1980's that the CFA was itself repopularized by Comlinear Corporation. To appreciate the evolution of the beast, it helps to look back to some early discrete transistor circuits.

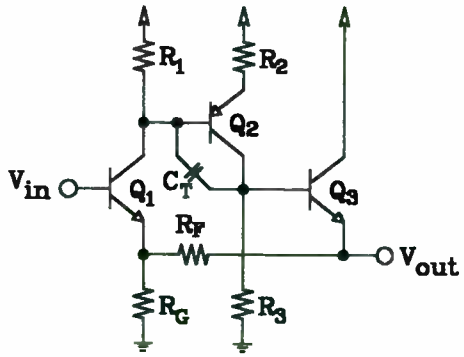


Fig.1: Three transistor amplifier

The three transistor amplifier of figure 1 is arranged in a series-shunt configuration. However, in order to analyze the amplifier, the circuit is rearranged as shown below in figure 2.

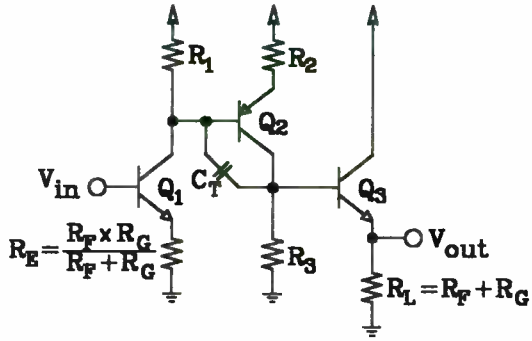


Fig. 2: Amplifier redrawn for analysis

The feedback network shows up in two places - a series network at the output and a parallel network at the emitter of the input transistor. This allows for open-loop analysis while keeping the effects of loading intact.

The loading of the output by the feedback network is generally not a problem. However, the gain of the first transistor stage is dependent on the values of the resistors in the feedback network. Thus the open-loop response will change with closed-loop gain, which could make frequency compensation an iterative chore.

Adding another transistor to buffer the input stage transistor from the feedback network can circumvent this difficulty. The discrete transistor circuit of figure 3 illustrates that this modification is the first step towards a voltage feedback amplifier.

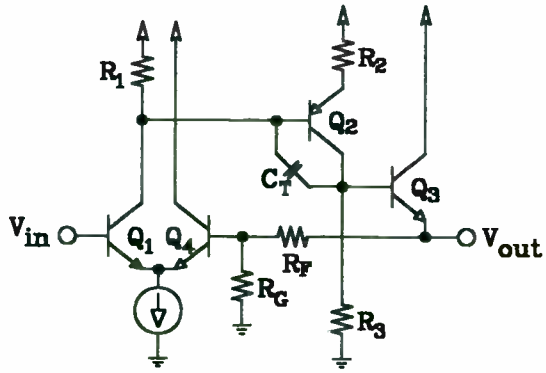


Fig. 3: Adding a buffer transistor

The added transistor presents a high impedance input to the feedback network. It also features the benefits of

a balanced input, such as low offset voltage and equal input bias currents.

Of greater significance is the fact that the dynamic emitter resistance of the added transistor is substituted for the parallel resistance of the feedback network in figure 2. The first stage gain, and consequently the open-loop gain, is no longer dependent on the feedback network. The process of frequency compensation has one less degree of freedom to be concerned with.

These two circuits illustrate the basic distinctions between current-feedback and voltage-feedback amplifiers. In both cases, the feedback network is connected to an inverting input node. In figure 1, the emitter presents a low impedance input while in figure 3 the base presents a high impedance input.

Needless to say, the three transistor amplifier of figure 1 can be considered the forbear for the CFA as it is known today. Figure 4 shows the amplifier connected to a mirror-image of itself, whose transistors have been converted to the opposite polarity type. Once the input transistors are buffered by emitter followers, the basic current-feedback architecture emerges.

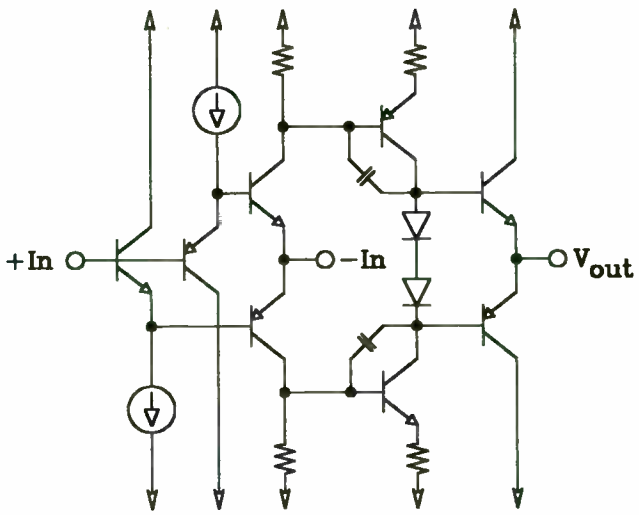


Fig. 4: Basic CFA topology

ANALYZING THE CFA

The study of the differential input, voltage-feedback amplifier is simplified with a technique known as "half-circuit analysis". This technique, illustrated in figure 5, recognizes that the symmetry of the circuit presents a shortcut whereby only half the signal path needs to be considered. The **NPN** current mirror, which provides double-ended to single-ended conversion, still maintains balance in the circuit because the second stage output voltage is determined by the current that flows into the high impedance presented by the collectors.

Inspection of figure 4 shows that the axis of symmetry for the CFA is centered horizontally. Therefore, the half-circuit used for analysis is the same circuit as presented in figure 1, ignoring the input emitter follower. However, as pointed out previously, the feedback network is closely intertwined with the analysis. Therefore, the circuit of figure 2 can be used for the analysis. The compensation capacitor, C_T, can be the intrinsic base-collector capacitor of Q2 or an extrinsic capacitor.

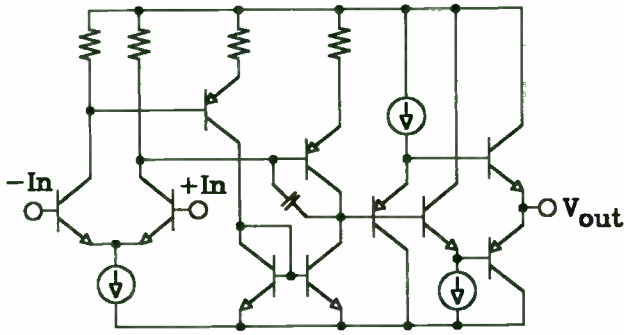


Fig. 5a: Basic VFA topology

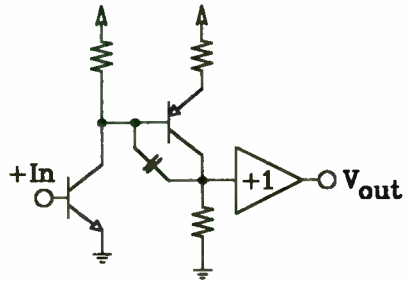


Fig. 5b: VFA half-circuit

The only real difference between figure 5b and figure 2 is the presence of the parallel combination of the feedback network resistors in the emitter of the CFA's input transistor. The analysis is very straightforward and the dc gain can be determined by inspection.

$$A_{Vdc} = \frac{R_1}{R_E} \cdot \frac{R_3}{R_2}$$

The open-loop pole can be approximated quite accurately as the interaction of the resistor, R_1, with the Miller multiplied capacitor, C_T.

$$\omega_p \cong \frac{1}{R_1\left(\frac{R_3}{R_2} \cdot C_T\right)}$$

This analysis presumes that the dynamic emitter resistance of Q1, r_{e1}, can be neglected ($R_E >> r_{e1}$) and that R_2 includes r_{e2}.

It would be convenient at this point to define the transresistance as:

$$R_T = \frac{R_1 \cdot R_3}{R_2}$$

Note that the transresistance has the dimensions of ohms and is determined solely by elements internal to, the amplifier. The previous equations can be rewritten more simply.

$$A_{Vdc} = \frac{R_T}{R_E} \text{ and } \omega_p = \frac{1}{R_T \cdot C_T}$$

Now the open-loop gain can be completely described by:

$$A_V = \frac{R_T}{R_E} \cdot \frac{1}{1 + j\omega R_T C_T}$$

In order to arrive at this equation, it was assumed that the feedback network was known. This is the crux of the issue - the open-loop **voltage** gain of a CFA requires knowledge of the feedback network.

Removing R_E, the feedback network term, from the equation for open-loop voltage gain yields a more general expression that describes the amplifier's open-loop performance in terms of its intrinsic characteristics. This equation would have units of ohms and would be better identified as a complex impedance, or transimpedance, Z_T:

$$Z_T = \frac{R_T}{1 + j\omega R_T C_T}$$

This is the true measure of performance for CFAs. It is now obvious why the amplifier is known as "current-feedback." The output voltage is responsive to a **current** at the low impedance inverting input node (the emitter of Q1) that interacts with the open-loop transimpedance, Z_T.

Furthermore, the open-loop response of the amplifier is completely described by the dc transresistance, R_T, and the compensation capacitor, C_T, which is called the transcapacitance. R_T is analogous to A_{OL}, the open-loop voltage gain of a VFA, and interacts with C_T to form the open-loop pole. This is graphically depicted in figure 6.

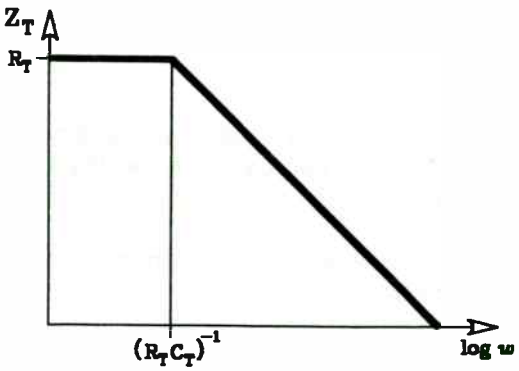


Fig. 6: Open-loop transimpedance

The ordinate axis has the dimension of ohms and is scaled logarithmically.

Having described the CFA with just two components suggests a simplified version of the half-circuit used for analysis. Figure 7 shows a convenient model that has all the essentials necessary for quick hand calculations. The inverting buffer preserves the sense of the signal as it is amplified by the Q2 stage in Figure 2.

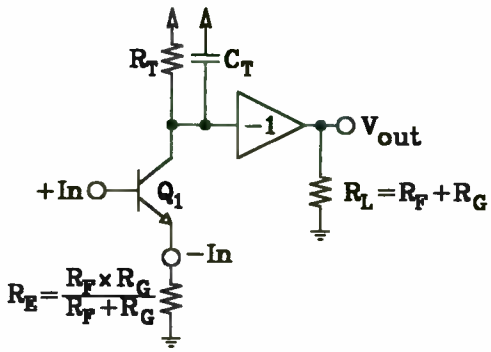


Fig. 7: CFA model for ac analysis

Comparisons with voltage-feedback amplifiers will inevitably be made when determining which op amp to use for an application. Presumably the closed-loop gain is known, which means that a feedback network can be established. Therefore, the open-loop voltage gain can be calculated for the CFA and a fair comparison with any VFA can be established.

Note that the analysis described here is based on a fairly simple current-feedback topology. Although the design of integrated circuit CFAs has become more sophisticated, the open-loop transimpedance approach (Z_T) is still valid.

CLOSED-LOOP PERFORMANCE

The closed-loop response of the CFA can be described by using classical analysis:

$$A_{CL} = \frac{A_V}{1 + A_V \cdot \beta} \text{ where } \beta = \frac{R_G}{R_F + R_G}$$

Substituting for A_V yields the following expression:

$$A_{CL} = \frac{\frac{R_T}{R_E} \cdot \frac{1}{1 + j\omega R_T C_T}}{1 + \frac{R_F}{\frac{R_T}{R_F} \cdot \frac{R_T}{1 + j\omega R_T C_T}} + j\omega R_F C_T} = \frac{\text{Open-loop gain}}{\text{Loop gain}}$$

The loop gain, of course, limits the accuracy of the closed-loop gain. Note that $R_T >> R_F$ (typically $R_T > 100K$ and $R_F < 5K$), therefore the equation can be easily simplified to:

$$A_{CL} = \frac{R_F + R_G}{R_F} \cdot \frac{1}{1 + j\omega R_F C_T}$$

The dc value of closed-loop gain is set by the feedback network while the closed-loop pole is determined by the interaction of the transcapacitance with the feedback resistor. This latter term is what gives the CFA its much touted characteristic of gain-independent bandwidth.

A closer look at the unsimplified equation for the closed-loop gain helps to clarify this property. The dc portion of open-loop gain in the numerator is modified by the parallel combination of the feedback network, which changes with desired closed-loop gain. As long as R_F is kept constant, the loop gain expression in the denominator does not vary, nor do any of the frequency dependent terms.

Figure 8 illustrates graphically that the open-loop gain curve slides vertically to keep the closed-loop intercept frequency constant.

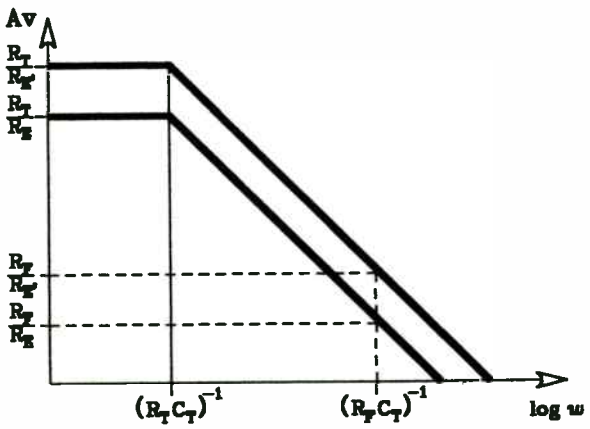


Fig. 8: Variation of open-loop gain

The closed-loop gain expressions have been expressed as a ratio of the feedback resistor to the equivalent feedback network. This can be verified algebraically as:

$$\frac{R_F}{R_E} = \frac{R_F}{\left(\frac{R_F \cdot R_G}{R_F + R_G}\right)} = \frac{R_F + R_G}{R_G}$$

Thus, the open-loop gain varies directly with the closed-loop gain for changes in R_E as long as R_F is kept constant.

NONIDEAL CONSIDERATIONS

The assumption that the r_e of Q1 can be neglected has limits. For ease of analysis, figure 6 has been redrawn to include it as an input resistance, R_{in} (figure 9). Note that R_{in} is internal to the CFA terminals.

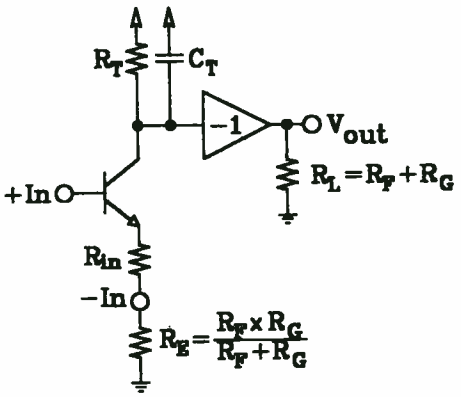


Fig. 9: CFA model modified for R_{in}

The open-loop gain equation can be modified by inspection while a new closed-loop gain equation can again be derived using the classical approach.

$$A_V = \frac{R_T}{R_E + R_{in}} \cdot \frac{1}{1 + j\omega R_T C_T}$$

$$A_{CL} = \frac{R_F + R_G}{R_G} \cdot \frac{1}{1 + j\omega \left(R_F + \frac{R_{in}}{\beta}\right) C_T}$$

R_{in} decreases the open-loop gain but not its corner frequency. On the other hand, R_{in} does not affect the dc closed-loop gain but does modify the intercept frequency. In practice, R_{in} includes more than just the dynamic emitter resistance - it also includes bulk resistances that are in series with the inverting input, as well as parasitic resistances external to the amplifier. Obviously, R_{in} should be as low as possible to get the maximum benefit from a CFA.

The modified equations lead to some practical generalizations when using CFAs. The first is that the open-loop gain has a theoretical maximum and this can be conveniently estimated as:

$$A_{V(max)} \cong \frac{R_T}{R_{in}} \cdot \frac{1}{1 + j\omega R_T C_T}$$

This is an ideal value that can never be realized since any feedback network will automatically reduce the open-loop gain. However, it is useful for estimating a CFA's merits against a particular VFA.

The second generalization is that the closed-loop bandwidth will become gain-bandwidth limited when

$$\frac{R_{in}}{\beta} \geq R_F \Leftrightarrow R_{in} \geq R_E$$

The latter expression makes use of the fact that the feedback factor, β, is a function of the feedback network resistors.

Once this limit has been reached, the CFA can be associated with a gain-bandwidth product, GBW.

$$GBW = \frac{1}{R_{in} C_T}$$

The graph in figure 10 shows an asymptotic approach to estimating a CFA's closed loop response.

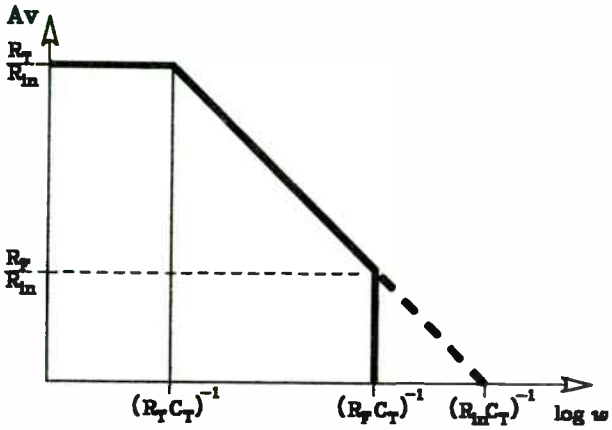


Fig. 10: CFA closed-loop performance

To be technically accurate, it should be pointed out that the inverting input is characterized by an impedance, Z_{in}, which does vary with frequency. Fortunately, the resistive portion, R_{in}, dominates over most of the CFA's useful bandwidth. At high frequency, the inverting input impedance increases, which only further degrades the closed-loop performance, although the extent of the increase is generally well under an order of magnitude.

FREQUENCY COMPENSATION

The analysis so far has centered on the gain versus frequency performance without taking into account any phase shift considerations. Excess phase plagues the CFA just as it does the VFA. The open-loop transimpedance curve of figure 6 depicts a single-pole response which would have only 90° of phase shift. Parasitic poles introduce additional phase shift to the open-loop phase response. Figure 11 displays the more complete open-loop transfer curves - both magnitude and phase.

Since the feedback network sets the open-loop gain for the CFA, it also sets the phase margin, Φ_M. This is the crucial factor that actually determines the selection of the feedback network resistors.

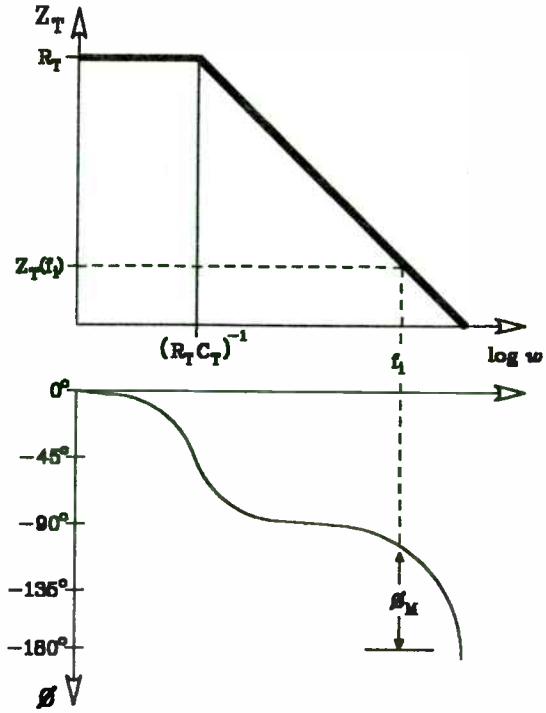


Fig. 11: CFA open-loop transfer curves

The significance of phase margin would benefit from a brief review of its properties. Phase margin, in a strictly literal sense, is measured at that frequency, f_U, where an amplifier's open-loop voltage gain has fallen to unity. It is the difference between the open-loop phase shift and -180°, where the amplifier would lose negative feedback and become unstable.

$$\Phi_M = \Phi(f_u) - (-180°)$$

The concept of phase margin is best illustrated by plotting unity gain frequency response curves as phase margin is varied (figure 12).

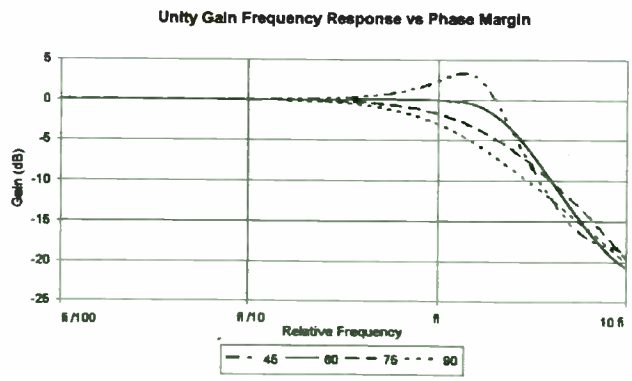


Fig. 12: Phase margin's effect on frequency response

As the plot shows, the optimum value for phase margin is 60°. This gives the desirable combination of broad bandwidth with flat frequency response. Note that an amplifier with 90° of phase margin, which implies a lack

of excess phase, has a -3dB bandwidth less than half of the optimum response.

A more general way of looking at this is to make the observation that the closed-loop response can be extended if the open-loop phase has fallen 120° at f_i, the frequency where the asymptote for closed-loop gain intersects the open-loop gain curve.

In VFAs, the phase margin is set by design and the user does not change it. There are a few amplifiers which allow access to the high impedance node to tailor compensation, but these are in the minority. In general, VFAs break out into two categories - compensated and decompensated.

The compensated amplifiers allow operation at unity gain but at the expense of bandwidth in higher gains. Decompensated, or undercompensated, amplifiers must be operated in gains greater than unity but have a higher gain-bandwidth product. In either case, the phase margin is predetermined.

As noted previously, phase margin for the CFA is set by the user via the feedback network. However, rather than use phase margin as the design criterion, higher performance can be attained by making use of the general observations regarding phase shift and bandwidth. In other words, guarantee that the open-loop phase has fallen 120° at f_i.

The mechanics are rather straightforward because, as illustrated in Figure 8, varying the feedback network causes a simple vertical translation of the open-loop gain curve. The open-loop pole does not move and so the attendant open-loop phase shift is unaffected. The excess phase shift is also insensitive to the feedback network change. Thus, selection of a desired phase shift automatically sets the intercept frequency.

Once the intercept frequency, f_i, is determined, so is the magnitude of transimpedance, $Z_T(f_i)$. This is depicted graphically in figure 11 by following the dashed lines up from the open-loop phase curve to the intersection with the open-loop transimpedance curve.

To realize the benefit of the -120° phase shift, the feedback network has to be selected so that the open-loop gain equals the closed-loop gain at f_i. A convenient way to visualize this problem is to concentrate on the essentials of the model in figure 9.

The CFA model can be simplified further by ignoring the inverting buffer and focusing on that portion of the circuit which provides gain. In figure 13 the CFA model has been reduced to an elementary transistor amplifier. The gain for this circuit is

$$|A_V| = \frac{|Z_T(f_i)|}{R_E + R_{in}}$$

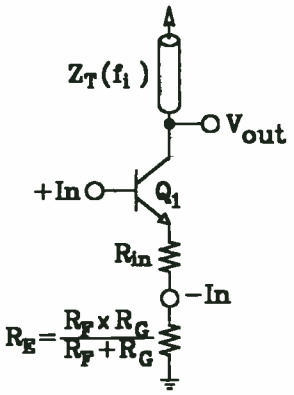


Fig. 13: Elementary amplifier

The goal, therefore, is to select the necessary feedback network so that A_V equals the desired closed-loop gain. Since Z_T has previously been defined as a complex impedance, direct substitution yields a closed form solution.

$$\frac{R_F + R_G}{R_G} = \frac{|Z_T(f_i)|}{R_{in} + R_E} = \frac{\left|\frac{R_T}{1 + j2\pi f_i R_T C_T}\right|}{R_{in} + R_E}$$

which can be reduced to a less bulky equation:

$$R_F \cong \frac{1}{2\pi f_i C_T} - \frac{R_{in}}{\beta}$$

Not surprisingly, this expression conforms to the plot of CFA closed-loop performance (figure 10). For low gains, the R_{in} term is negligible and R_F is set by f_i. As closed-loop gain increases and R_{in}/β can no longer be neglected, R_F should be adjusted according to the equation to maintain optimum performance. When R_F approaches zero, the CFA is becoming gain-bandwidth limited and the intercept frequency must be lowered.

MODEL REPRESENTATION

The single transistor model of figure 9 is a satisfactory vehicle to provide intuitive insight. It is by no means an accurate representation of the CFA but offers a good visual aid for the user.

A more generally accepted model for the CFA is depicted in figure 14. This model is a very faithful rendition of the CFA from a block diagram standpoint. It can accurately account for the bipolar input and output swings that are possible with the CFA's complementary symmetry.

Comparing it to figure 4, it is readily apparent that the unity gain buffer at the input is an accurate portrayal of the input stage between the input pins. The finite input resistance, R_{in}, is included for completeness.

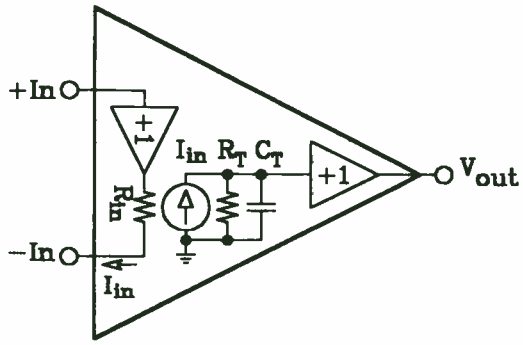


Fig. 14: Block diagram CFA model

The current-controlled current source translates the current from the inverting input to the open-loop transimpedance, again composed of R_T and C_T. The unity-gain buffer provides a low impedance source to the external load.

Either of the models is sufficient to appreciate the CFA and its performance features. Figure 9 bears a strong resemblance to the ancestral antecedent of the CFA while the latter is more readily adaptable to generating a SPICE macromodel.

Other properties of the CFA are apparent when studying these models. The slew rate is limited by the current available to charge the transcapacitance. Decreasing $(R_{in}+R_E)$ will certainly benefit slew performance. Minimizing C_T will increase slew rate as well as the small-signal performance.

Potential for trouble exists when parasitic capacitance is present at the inverting input. This parasitic capacitance can be the result of poor layout techniques, inappropriate use of a socket or even the wrong package. If C_P is the lumped parasitic capacitor, the open-loop gain will become:

$$A_V = \frac{R_T}{R_E + R_{in}} \cdot \frac{1 + j\omega R_E C_P}{(1 + j\omega R_T C_T)\left(1 + j\omega \frac{R_E \cdot R_{in}}{R_E + R_{in}} C_P\right)}$$

This expression has added a zero and a pole to the transfer function. The zero will always occur before the pole and can be the source of trouble in some cases. If instability arises because of C_P, move the **closed-loop** pole to a lower frequency by adjusting the feedback network.

To model excess phase, the addition of a delay line can be more expedient than trying to add multiple poles and zeroes to the open-loop transimpedance. The modified transfer function is still quite compact.

$$Z_T = \frac{R_T}{1 + j\omega R_T C_T} \cdot e^{-j\omega T_D}$$

The exponential adds phase shift without affecting magnitude. A reasonable technique is to use the phase shift at the highest intercept frequency the circuit is expected to encounter.

$$T_D = \frac{1}{2\pi f_i \cdot \frac{\Phi(f_i) - 90°}{360°}}$$

Subtracting 90° from the open-loop phase is, of course, to remove the phase shift due to the open-loop pole.

DATA SHEET SPECIFICATIONS

The open-loop transimpedance terms, R_T and C_T, and the input resistance, R_{in}, have already been identified as necessary features to describe a CFA. Additionally, the open-loop transimpedance and phase versus frequency curves should be provided as well.

The block diagram presentation of figure 14 suggests the other specifications that should not be overlooked. The presence of a buffer between the noninverting and inverting inputs of the CFA guarantees that the input characteristics will not match. This is the main difference between the VFA and the CFA data sheets.

The VFA data sheet typically specifies the power supply and common-mode rejection for the offset voltage only. The input bias currents are also subject to disturbances from these sources but good VFA design encourages matching impedances at the inputs to mask the effects.

The CFA does not have the privilege of bias current match, so the same effects that are specified for the offset voltage need to be measured for the two input nodes. In particular, the inverting input, which is the true signal input is often the biggest source of error. It is not uncommon to see a CFA constrained to operate in an inverting gain configuration to circumvent common-mode effects.

It is not very common practice to specify power supply rejection for each supply separately but, for the CFA, it is essential. The complementary devices, NPN and PNP, should not be expected to match each other closely and usually the PNPs are the weaker. PSR measured with tracking supplies typically tend to partially cancel the errors. Real world applications usually rely on independent positive and negative voltage regulators.

The table below is for a medium performance CFA and exemplifies the amount of detail that should be provided.

INPUT OFFSET VOLTAGE		
Initial	5	mV
vs Temperature	8	uV/°C
vs Common-mode	60	dB
vs Supply (tracking)	85	dB
vs Supply (non-tracking)	60	dB
+INPUT BIAS CURRENT		
Initial	5	uA
vs Temperature	30	nA/°C
vs Common-mode	200	nA/V
vs Supply (tracking)	50	nA/V
vs Supply (non-tracking)	150	nA/V
-INPUT BIAS CURRENT		
Initial	25	uA
vs Temperature	300	nA/°C
vs Common-mode	200	nA/V
vs Supply (tracking)	300	nA/V
vs Supply (non-tracking)	1500	nA/V
INPUT IMPEDANCE		
+Input	5M//2	Ω//pF
-Input	30//2	Ω//pF
OPEN-LOOP TRANSIMPEDANCE		
Transresistance	440	kΩ
Transcapacitance	1.8	pF
OUTPUT CHARACTERISTICS		
Voltage	12	V
Current	150	mA
Output resistance, open-loop	70	Ω

(Source: BB OPA603 data sheet)

SPICE SIMULATION

The combination of declining hardware costs with increasing computing horsepower has made circuit simulation a required part of the design cycle. This has forced the op amp vendor to supply the macromodels for his product offering.

These simulation tools have been offered in varying degrees of complexity, from the simple Boyle model to simplified circuit models, which utilize full transistor models in the signal path. There has been a growing consensus that this latter approach is necessary for the high bandwidth amplifiers.

There can be no doubt that having these models available helps to fill in the gaps from incomplete data sheets. Although the models may not necessarily be configured for worst case process extremes, there may be some performance peculiarities that can be discovered through their use. The pitfall to be aware of is that even the simplified circuit models generally idealize the biasing circuitry, which may mask some second order PSR and CMR effects.

Figure 15 shows two alternative simulation schemes. In figure 15a, the CFA is driven open-loop to measure the open-loop transimpedance and input resistance. This requires two separate simulations. The first uses a voltage-controlled current source to find the dc value of inverting input current to servo the output to zero. The second pass is the ac simulation to actually measure the transimpedance.

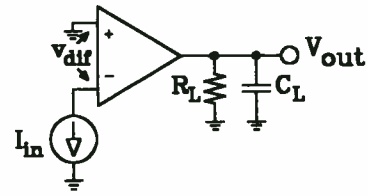


Fig. 15a: Open-loop simulation

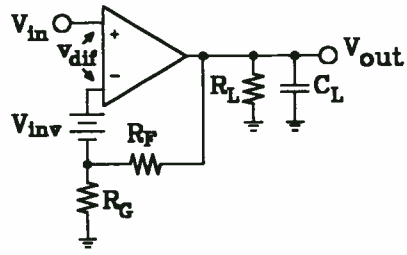


Fig. 15b: In circuit measurement

Figure 15b uses a zero volt battery to measure the inverting input current while the op amp is in a closed-loop configuration. This measures an effective transimpedance that includes the common-mode effect.

The circuit of figure 15a was simulated with the following listing:

```
* CURRENT-FEEDBACK OPEN-LOOP SIMULATION *
* file: CFA-OL.CIR
***** Simulation Commands *****
.options noecho nomod numdgt=8
.op
.ac dec 20 10 200meg
.probe
***** Library Files *****
.lib bb-updat.lib
***** Circuit Listing *****
vp 7 0 15
vm 4 0 -15
*ginv 2 0 6 0 -1
inv 2 0 dc -38.3pa ac 1
x603 0 2 7 4 6 opa603
rl 6 0 100k
.end
```

Figure 16a is the plot of input resistance as measured by dividing the ac voltage by the ac current. Note that for the useful frequency range of the amplifier (roughly 100MHz), R_{in} varies less than 10Ω. The open-loop transimpedance is displayed in figure 16b. Here the magnitude has fallen from a dc value of 790kΩ to 1.5kΩ at 51.6MHz, which is where the open-loop phase has fallen to -120°.

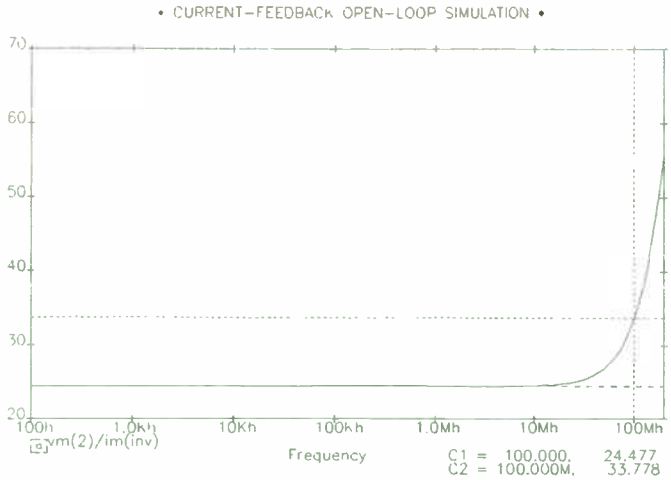


Fig. 16a: Measuring the inverting input impedance

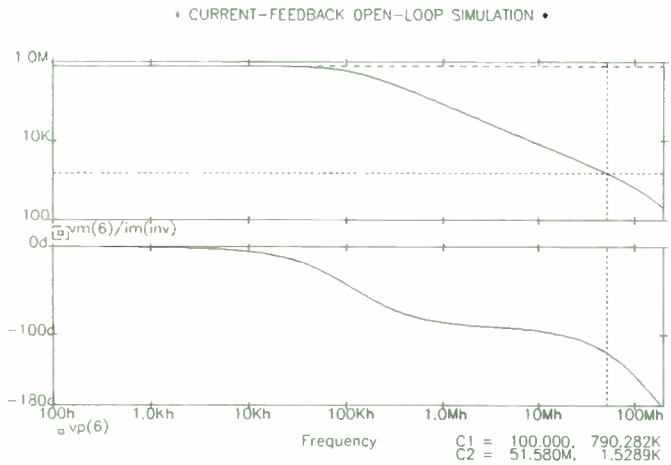


Fig. 16b: Measuring open-loop transimpedance

The circuit of figure 15b was simulated with the following listing:

```
* CURRENT-FEEDBACK CLOSED-LOOP SIMULATION *
* file: CFA-CL.CIR
***** Simulation Commands *****
.options noecho nomod
.ac dec 20 1000 200meg
.probe
***** Library Files *****
.lib bb-updat.lib
***** Circuit Listing *****
vp 7 0 15
vm 4 0 -15
vin 3 0 dc 0 ac 1
x603 3 inv 7 4 6 opa603
vinv inv 2 dc 0
rf 6 2 1450
rg 2 0 1450
.end
```

The plot in figure 17 shows the intersection of the open-loop gain curve with the closed-loop gain asymptote which occurs at 45.7MHz. The open-loop phase has the value of -120° at this frequency and the broadbanding of the closed-loop gain is quite evident. Note the technique used to generate the open-loop gain curve.

The equation relies on the calculation of open-loop transimpedance (via the current in the battery) which is divided by the sum of the equivalent feedback network plus the input resistance.

$$|A_V| = \frac{|Z_T(f_i)|}{R_E + R_{in}} = \frac{\frac{vm(output)}{im(battery)}}{R_E + R_{in}}$$

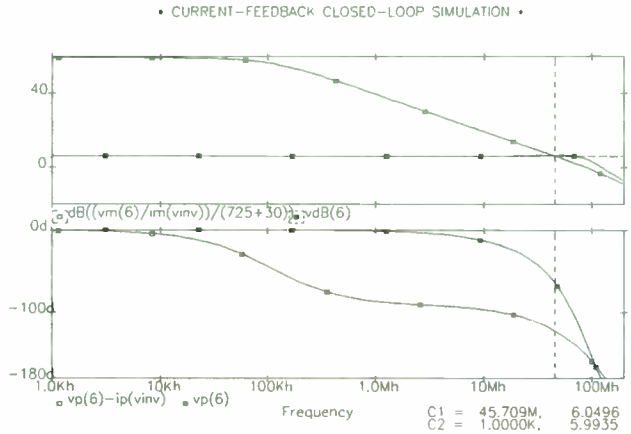


Fig. 17: Slope intercept curves for CFA circuit

MEASUREMENT CIRCUITS

If companies could just ship simulation files to their customers, life would be so easy. Sooner or later, a reality check has to be made. The following circuits have been proven to be quite reliable for measuring the CFA performance parameters.

The low impedance of the inverting input node presents a special problem for the test engineer. Conventional op amp test circuits cannot easily separate the individual parameter variations. The most logical solution is to test the CFA with a current mode test circuit.

Figure 18 shows the basic current pump topology used in the dc test circuit. It consists of an op amp, a P-channel MOSFET and a unique current reference circuit which includes two very accurate current sources and a high precision current mirror.

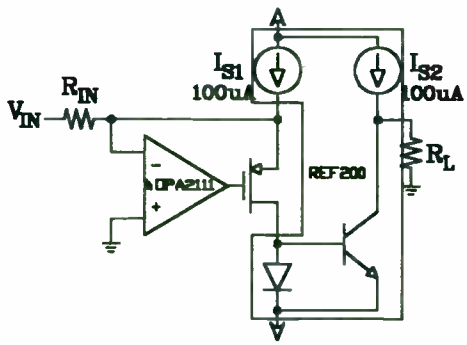


Fig. 18: Current pump topology

The high gain of the JFET input op amp (VFA) constrains its inverting input to stay at null ground by controlling the current flowing through the MOSFET. If V_{IN} is positive, a current equal to V_{IN}/R_{IN} is shunted to the current mirror input. If V_{IN} is negative, a matching V_{IN}/R_{IN} is provided by the 100uA current source, I_{S1}, and the input to the current mirror decreases. This is an inverting current pump, a positive voltage causes the output to sink current and a negative input causes the output to source current.

The full test circuit is shown in figure 19. The input offset voltage of the DUT is measured directly by the instrumentation amplifier, A1. The RC filters minimize noise and protect the inputs of A1 from overload transients.

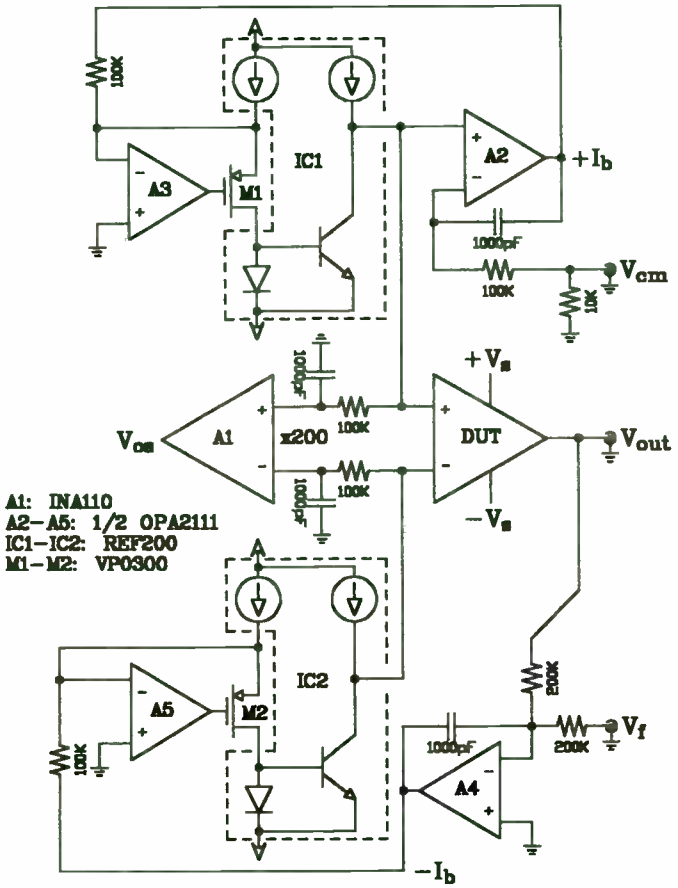


Fig. 19: Current mode CFA test circuit

Amplifier A2 maintains the common-mode bias by forcing the current pump (A3, M1, IC1) to keep the noninverting input of the DUT equal to the input, V_{CM}. The output of A2 driving the 100kΩ input resistor to the current pump is a measure of $+I_b$.

Amplifier A4 constrains the DUT output to be the negative of the input voltage, V_f, by forcing the current pump (A5, M2, IC2) to drive the low impedance inverting input. The amount of inverting current drive is reflected by the output of A4.

All dc parameters, including R_T and R_{in}, can be measured independently and directly. When adapted to a measurement card for the HP Semiconductor Analyzer, the test parameters can be displayed as slopes to determine the limits of linearity.

Figure 20 details an open-loop transimpedance test circuit which, when mated with a network analyzer, will provide the open-loop frequency response curves.

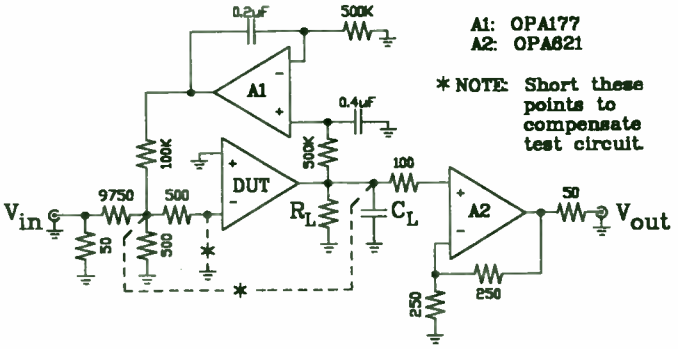


Fig. 20: Open-loop frequency response test circuit

The input ladder network divides the input by 20,000 to provide a low current level signal to the inverting input of the DUT. The 500Ω value for the input resistor dominates the small but finite input resistance of the CFA. The A1 integrator servos the output to zero by sensing the DUT output and feeding a small current back to the input. A2 buffers the DUT output and drives the 50Ω input of the network analyzer.

The only caveat is to take into account the gain and phase rolloff of A2. Automated network analyzers allow for compensation by storing an "offset" sweep which is subtracted from the actual signal sweep.

Although the network analyzer will scale the output in dB, the transimpedance can be determined by using the following equation:

$$Z_T = 500 \cdot \log^{-1}\left(\frac{\text{dB magnitude}}{20}\right)$$

The transcapacitance can be found by extrapolating the open-loop pole.

CONCLUSION

CFAs are not difficult to comprehend and work with if the basic relationships between R_T, C_T, R_{in} and open-loop phase are kept in mind. The lack of balanced input nodes require extra care be taken with applications requiring dc accuracy. Simulation is a wonderful tool for the early design stages but only actual measurements will grant peace of mind.

The SLAM: A New Ultra-linear Power FET Module Concept for HF Applications.

Adrian I. Cogan
Lee B. Max (*)
Al Rosenzweig

MicroWave Technology, Inc
4268 Solar Way
Fremont, CA 9453

Introduction

This paper describes the application of a new family of low cost silicon FET linear Class-A HF power gain blocks. *

These gain blocks, called 'SST Linear Amplifier Modules', or 'SLAMs' use a high frequency power FET; the 'Solid State Triode™', or 'SST™'. The SLAM modules are self-biased, and exhibit excellent linearity and thermal stability. With these modules, Class A broadband amplifiers covering the frequency range of DC through 100 MHz can be easily constructed with power outputs ranging from 10 Watts to 200 Watts or more.

SLAMs can easily be power combined to deliver hundreds of watts of linear HF power. Several 10 W through 200 W class-A power amplifiers, operating from 2 MHz through 32 MHz and 10 MHz through 100 MHz are described. The actual construction of the SLAM circuits is also described.

Inside the SLAM: The SST and Self-biased SST Operation.

The active element in the SLAM building block is a high frequency Silicon FET chip called the Solid State Triode, or SST. The SST name comes from the device's I-V characteristics which are similar to those of a vacuum tube triode (Figure 1).

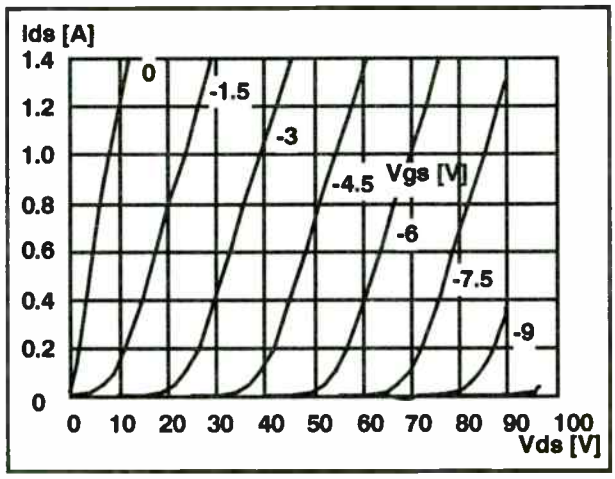


Figure 1: SST I-V Characteristics.

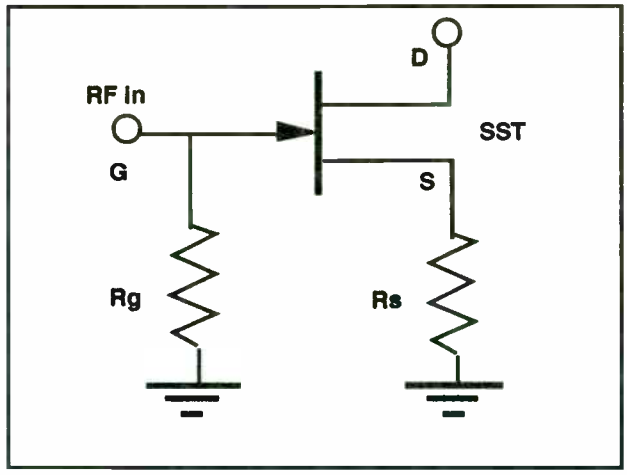


Figure 2. SLAM Internal Biasing Circuit.

The SST is a Junction FET and requires the same gate and drain bias polarity as a GaAs MESFET or a vacuum triode. Such a bias polarity (positive Vds and negative Vgs) allows one to set the SST operating point by using the self-biasing circuit of Figure 2.

* Mr. Max is an independent consultant

The voltage drop produced by the source current across the source resistor sets the SST operating point by supplying the gate voltage through the gate-to-ground resistor. This biasing scheme will set a Class A operating point and, if the source resistor is not bypassed with a low reactance capacitor, the resulting negative feedback will improve the circuit linearity. Added advantages are flatter frequency response, higher input impedance, and improved stability. The resulting I-V characteristics of a self-biased SST are shown in Figure 3. Notice the current saturation trend, despite the SST triode-like characteristics.

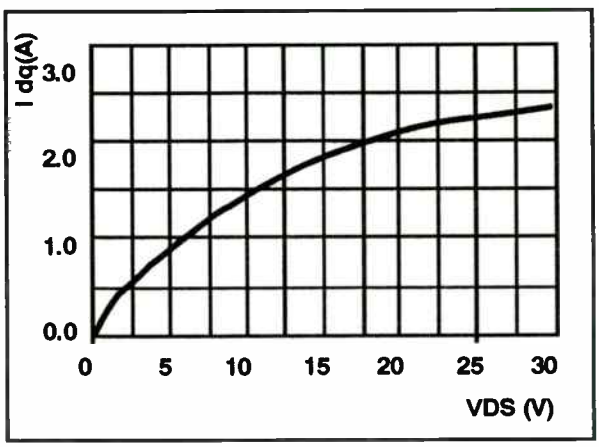


Figure 3. SLAM0111: Bias Current vs. Vds.

The SST exhibits excellent linear performance, as illustrated in Figure 4, where a 50 W SST third-order intermodulation product ('IMD3') is compared to similar power MOSFETs and bipolar transistors. This data was taken at 215 MHz without the self-biasing feature. The self-biasing circuit improves the SST linearity even further.

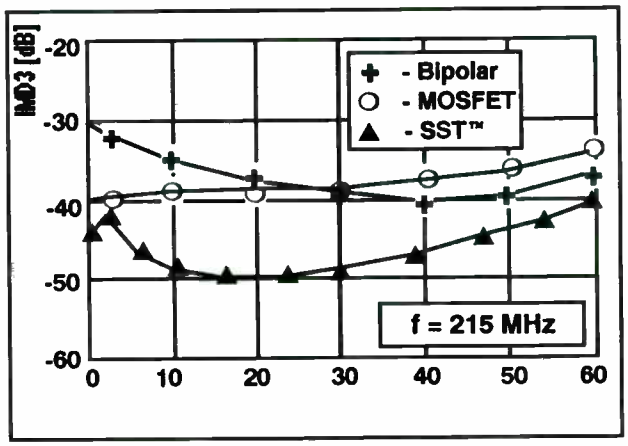
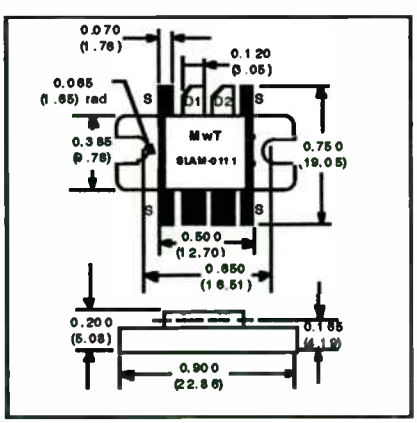


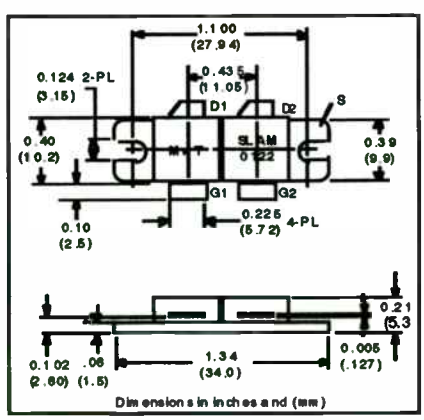
Figure 4. Typical Class AB Two Tones IMD3: 50 W BJT, MOSFET, and SST

By assembling the SST chip and the gate and source resistors inside a conventional RF transistor package, the result is a SST Linear Amplifier Module, or a 'SLAM'. The SLAM uses only a single power supply and performance remains constant over a wide range of supply voltage.

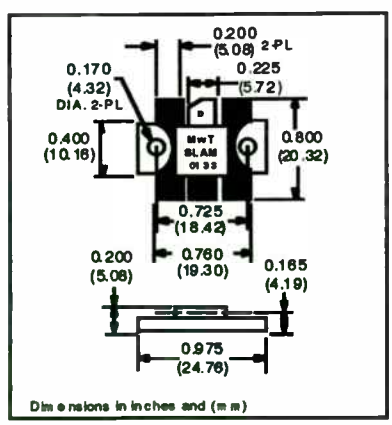
Presently, MwT supplies three power SLAMs: a single-ended unit, the SLAM-0133 rated for 10 W; and two push-pull SLAMs, the SLAM-0111, rated for 25 W, and the SLAM-0122, rated for 50 W. All three devices are designed to operate from DC through 32 MHz. SLAMs are supplied assembled in industry standard single-ended or push-pull packages (Figure 5).



SLAM-0111



SLAM-0122



SLAM-0133

Figure 5. SLAM Packages.

The SLAM and SLAM Circuit Building Blocks

The values of the SLAM internal components are selected to minimize or even eliminate the need for impedance matching components when operating from a 50 Ω source and load impedance.

The SST FETs in combination with the internal self-biasing circuit provide the user with a thermally compensated circuit component. The SLAM power gain has a thermal derating factor of less than 0.01 dB/°C and the SLAM bias point is virtually constant with temperature.

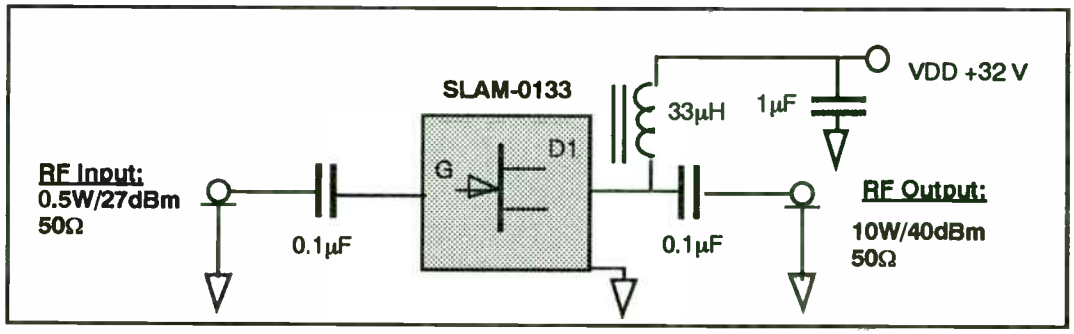


Figure 6. SLAM-0133, 2 MHz - 32 MHz Driver-Amplifier.

The simplest SLAM circuit is shown in Figure 6, where a single-ended unit is used as a 10 W HF driver. The operating bandwidth is 2 MHz to 32 MHz. Notice the simplicity of this amplifier which requires, in addition to the SLAM, only two DC-blocking capacitors and one RF choke. Figure 7 illustrates the frequency response of this amplifier.

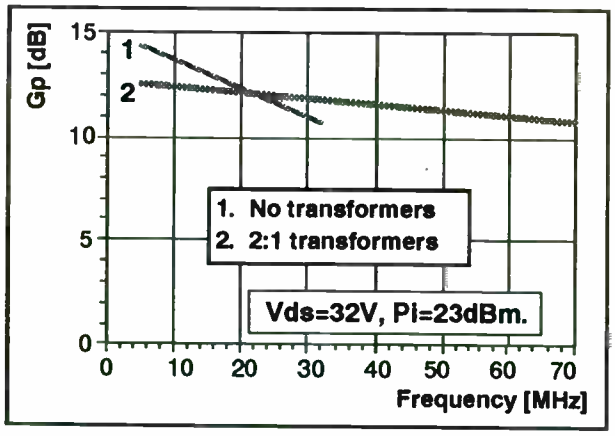


Figure 7. Typical Gp vs. Freq.

By adding inexpensive 2:1 input and output transformers, the frequency roll-off is reduced and the bandwidth expanded. Such a transformer can be implemented using 50 Ω coaxial cables (Figure 8).

The bandwidth can be further widened to 100 MHz by using additional impedance matching components, but this will increase the circuit complexity.

The third order intermodulation products (IMD3) and the harmonic contents (f2, f3) vs. the input power level for this amplifier are shown in Figures 9 and 10, respectively. The third order intercept point ('TOIP', or 'IP3') is +52 dBm, more than 10 dB above the P-1dB level.

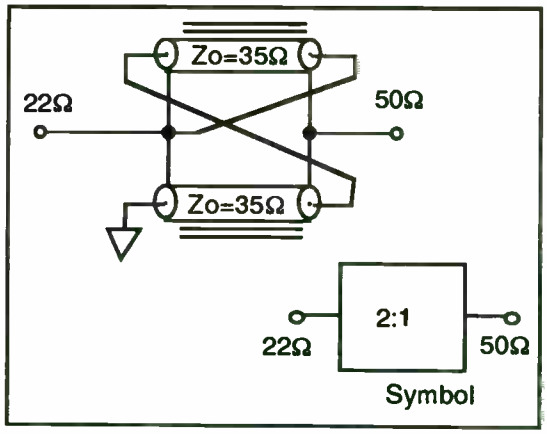


Figure 8. 2:1 Impedance Transformer.

Figure 11 shows the input VSWR versus frequency for this simple SLAM amplifier while power gain vs. supply voltage is shown in Figure 12.

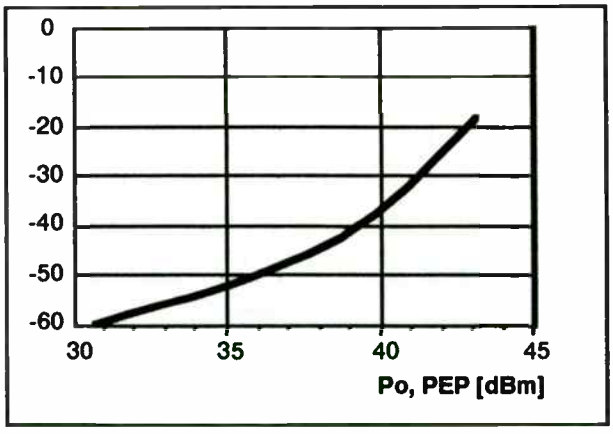


Figure 9. Typical IMD3 vs Power Output

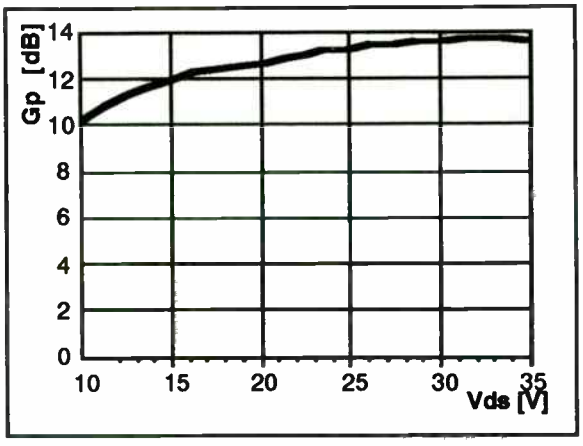


Figure 12. Typical Gp vs. Vds, f=8MHz

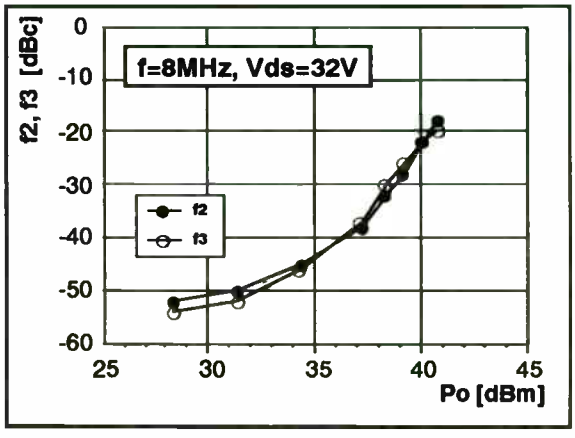


Figure 10. Typical Harmonics Content

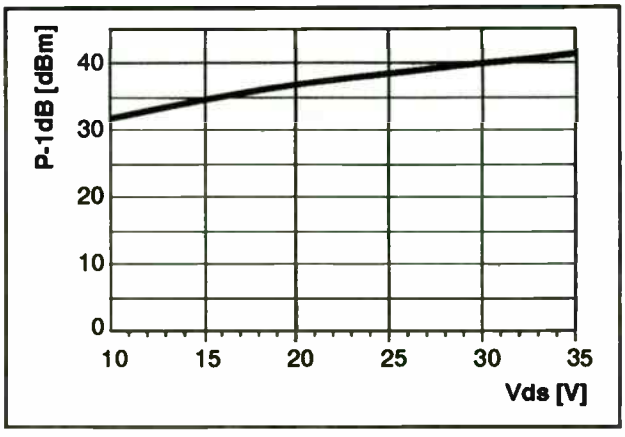


Figure 13. Typical P-1dB vs. Vds

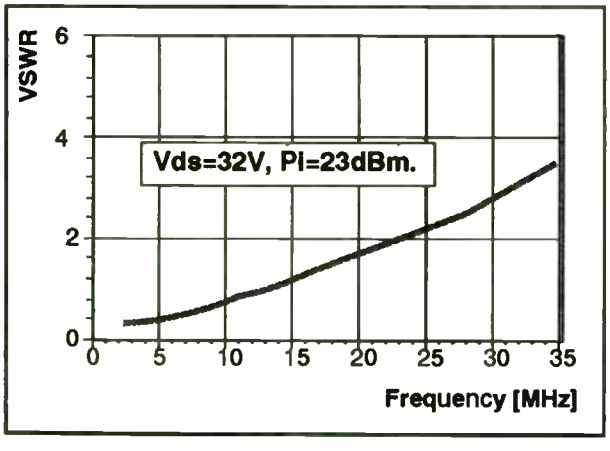


Figure 11. Typical Input VSWR vs. Frequency.

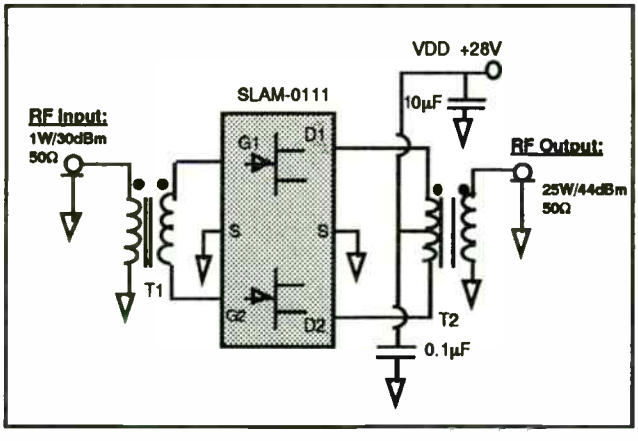


Figure 14. 2MHz - 32MHz SLAM Amplifier Circuit.

Figure 13 plots the one-dB compression point (P-1dB) dependence upon supply voltage. As shown, at the recommended operating point, the gain remains constant over a wide supply voltage variation.

A 25 W push-pull SLAM-0111 amplifier circuit is shown in Figure 14. The amplifier operates over the 2 MHz - 32 MHz HF band.

This configuration uses wire wound transformers. Notice again the simplicity of the circuit.

The wire wound 1:1 input transformer is made by winding 2 to 5 turns of #30 AWG wire on a ferrite balun core. The 1:1 output/bias trans-former uses 3 turns, center tapped #26 AWG wire on a balun core. The

above figures cor-respond to a core with a μ = 2000. The actual number of turns is determined by the ferrite actual μ and the desired amplifier bandwidth. This SLAM amplifier operates at a 25 W nominal output power and exhibits a TOIP of +55 dBm. Figure 15 shows the IMD3 data at five frequencies, while P-1dB and power gain vs. supply voltage is plotted in Figure 16.

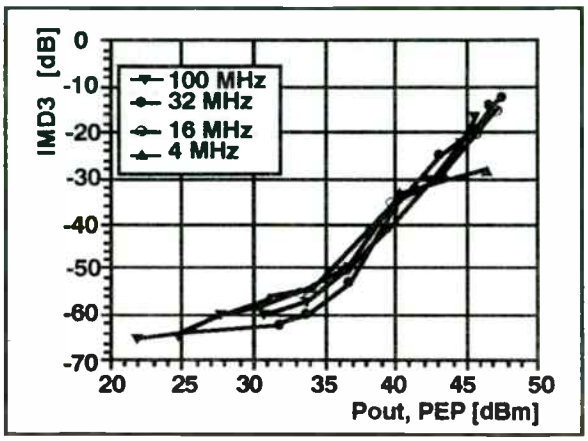


Figure 15. 25W SLAM:0111 IMD3 vs. f, Vds = 28 V.

Again, notice the reduced sensitivity to the supply voltage when operating this amplifier at the recommended voltage. One can broaden the operating bandwidth of this amplifier to 1 MHz through 100 MHz by using the circuit of Figure 17. The maximum linear output power will, however, be reduced to 20 W at the high frequency end of the band.

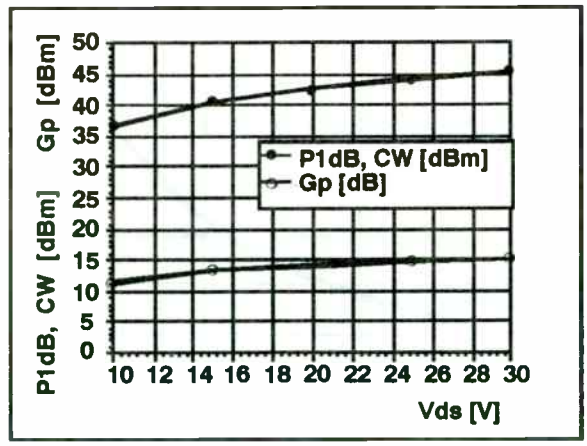


Figure 16. 25W SLAM: P-1dB, Gp @ 16 MHz

The wire wound transformers are replaced by coaxial baluns and a coaxial biasing transformer has been added between the two drain terminals.

The input transformer can be implemented by winding three to eight turns of RG-174U, 50 Ω coaxial cable over a ferrite toroid core with a μ ≥ 100. The bias transformer uses 4 turns of the same 50 Ω coax and ferrite core type. The frequency response of this SLAM amplifier is shown in Figure 18.

One advantage of the SLAM concept is its modularity. By using a small number of impedance transformers, power splitters, and bias transformers, one can power-combine SLAMs to almost any practical power level. Several such 'modular' SLAM amplifiers will be described below.

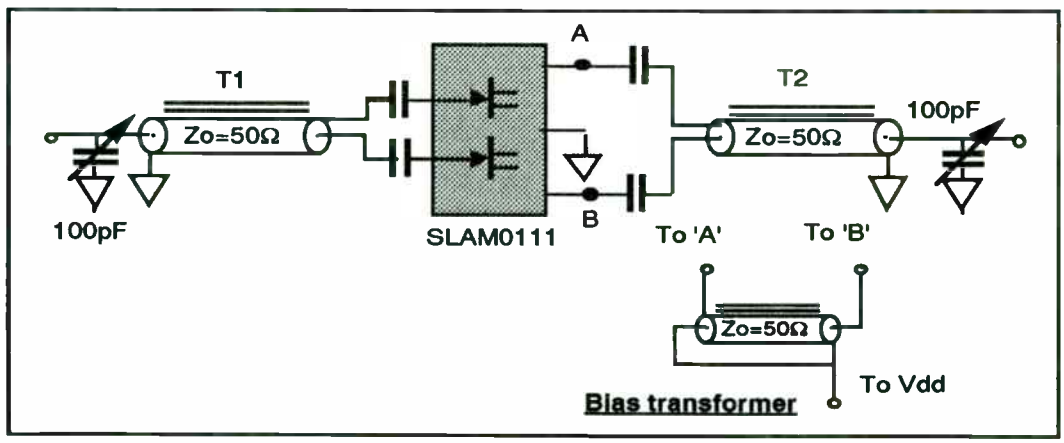
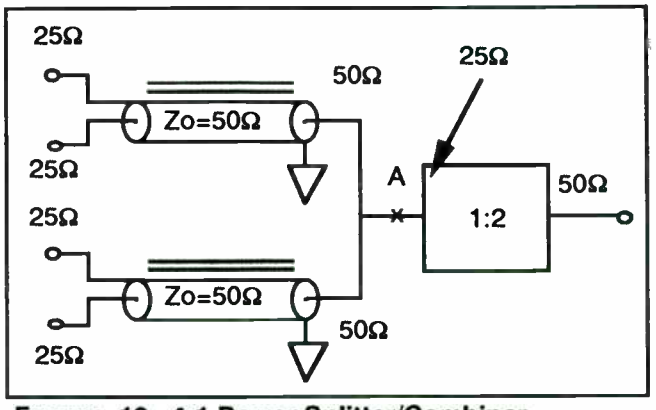


Figure 17. 20W, 1MHz - 100MHz SLAM0111 Amplifier.

Two coaxial baluns and one 2:1 transformer (Figure 8) can be combined in the configuration shown in Figure 19, to yield a two-way power splitter/combiner with Zin = Zout = 50 Ω. As it will be shown below, this

combiner is used to assemble 50 W and 100 W push-pull SLAM amplifiers.



Fugure 19. 4:1 Power Splitter/Combiner with (optional) improved isolation.

Another useful coaxial cable power combiner with Zin = Zout = 50 Ω is the four-way unit shown in Figure 20. Such combiners are used to assemble 100 W and 200 W power SLAM amplifiers.

An isolation transformer, such as the one illustrated in Figure 21, can be inserted at the point marked 'A' in the Figure 19 two-way combiner. This option will provide improved load isolation for the high power SLAM amplifiers.

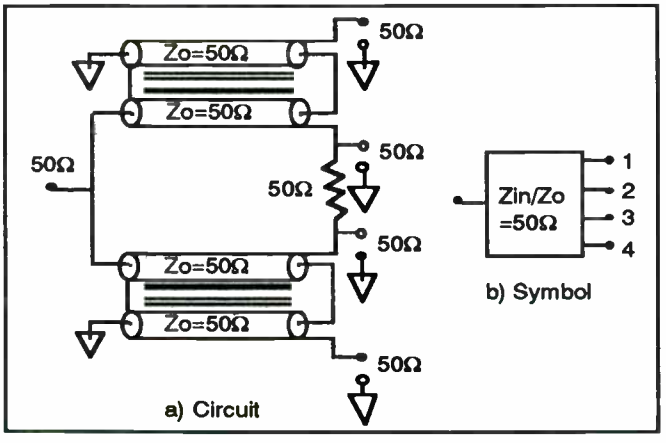


Figure 20. 50Ω/50Ω, 1:4 Power Splitter/Combiner.

Class A Linear, Power Amplifiers.

Using the impedance transformers and splitter/combiners described above, individual SLAM packages can be combined to achieve higher power levels over the HF frequency range.

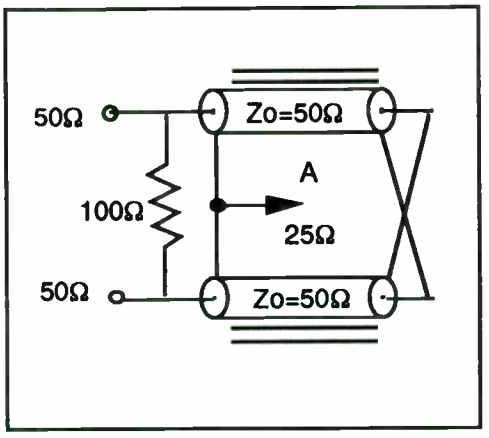


Figure 21. Isolation Transformer.

One simple combination was shown in Figure 17 where two coaxial cable baluns and one coaxial cable bias transformer are used to assemble a push-pull amplifier. The bias transformer is connected between the points marked 'A' and 'B', which represent the SST drains.

Over the 2 MHz - 32 MHz frequency range, the amplifier will yield 25 W of linear power if the SLAM-0111 is used, and 50 W with the SLAM-0122. The recommended bias voltages are 28 V for the SLAM-0111 and SLAM-0133, and 40 V for the SLAM-0122.

By combining two of the above SLAM-0111 amplifiers with two 50 Ω coaxial baluns, one will achieve 10 MHz through 100 MHz operation, with a 40 W output power (Figure 22). The inner baluns use 25 Ω coaxial cable and the same number of turn as the outer ones.

Biasing transformers are not illustrated, but they must be connected to the SLAM drains as in the Figure 16 example.

The frequency response of the SLAM-0111 version of this amplifier, which has a broadband gain of 12 dB, is shown in Figure 23.

The 2:1 transformers and 50 Ω coaxial baluns can be used as in Figure 24 to build 50 W SLAM-0111, or 100 W SLAM-0122 amplifiers. Two bias transformers will supply DC power to each of the two SLAM drain pairs. This amplifier operates over the 2 MHz - 32 MHz HF range.

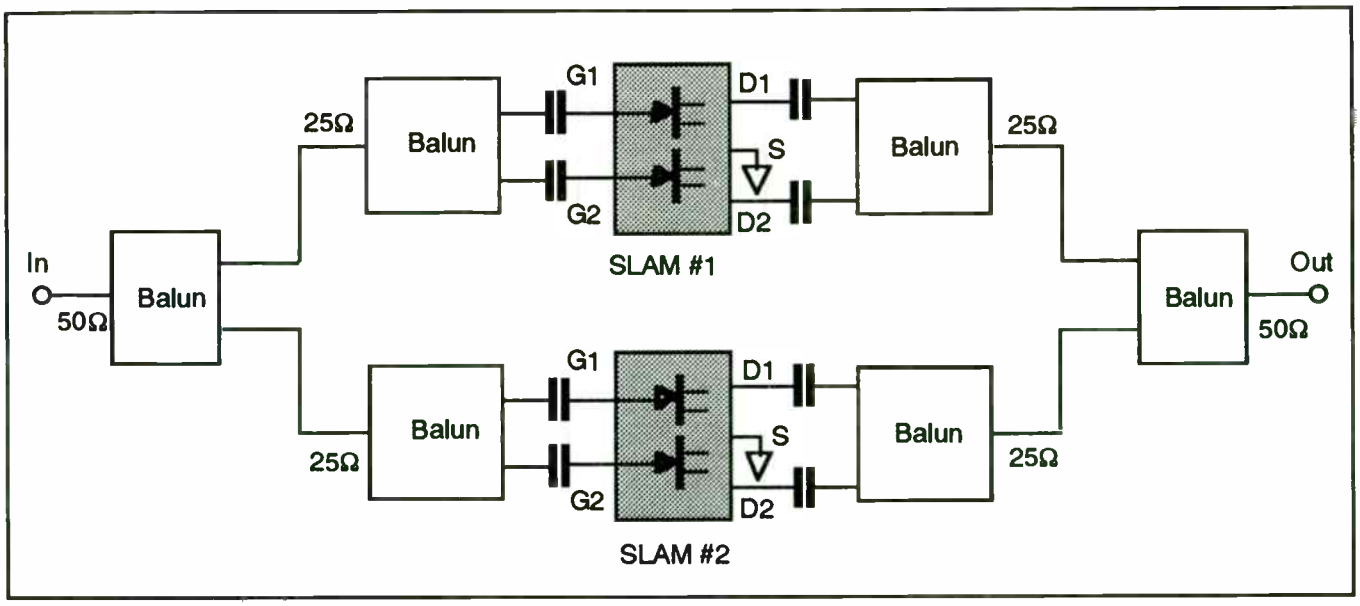


Figure 22. 40W, 10MHz - 100MHz, SLAM0111 or 75W 1MHz - 50MHz, SLAM0122 Amplifier.

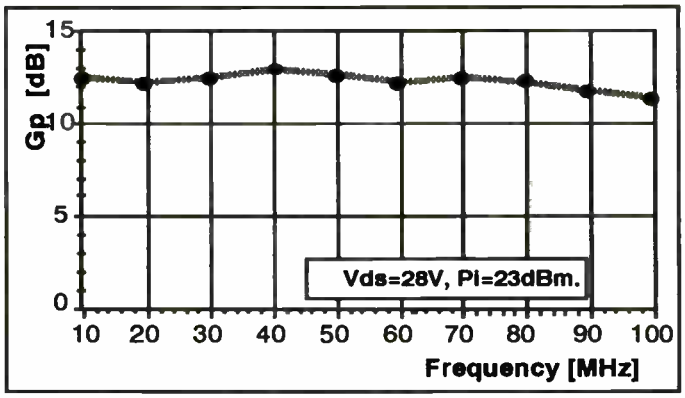


Figure 23. Broadband 50W 2xSLAM0111 Amplifier: Typical Gp vs. Frequency.

By using the coaxial transformer in Figures 19 and 21, improved isolation can be achieved

One can use four SLAMs as shown in the example in Figure 25, and implement 200 W SLAM-0122, or 100 W SLAM-0111 HF amplifiers. The four-to-one power combiners of Figure 20 are used in this example.

Conclusions

Linear, Class A, HF amplifiers with power outputs of from 10 W to over 200 W can be fabricated with far fewer circuit elements using internally matched, self-biased gain blocks that employ Solid-State Triode JFETs. Only simple transformers are required for combining push-pull elements or to combine several SLAMs into higher power amplifiers.

Acknowledgements.

The authors acknowledge the critical contribution made at MwT by N. Thornton and P. Pacada, and the support and guidance provided by M. Omori.

Bibliography

A. Cogan, "SST for Microwave Applications", RF Expo East Proceedings, October 1989, Atlantic City, NJ.

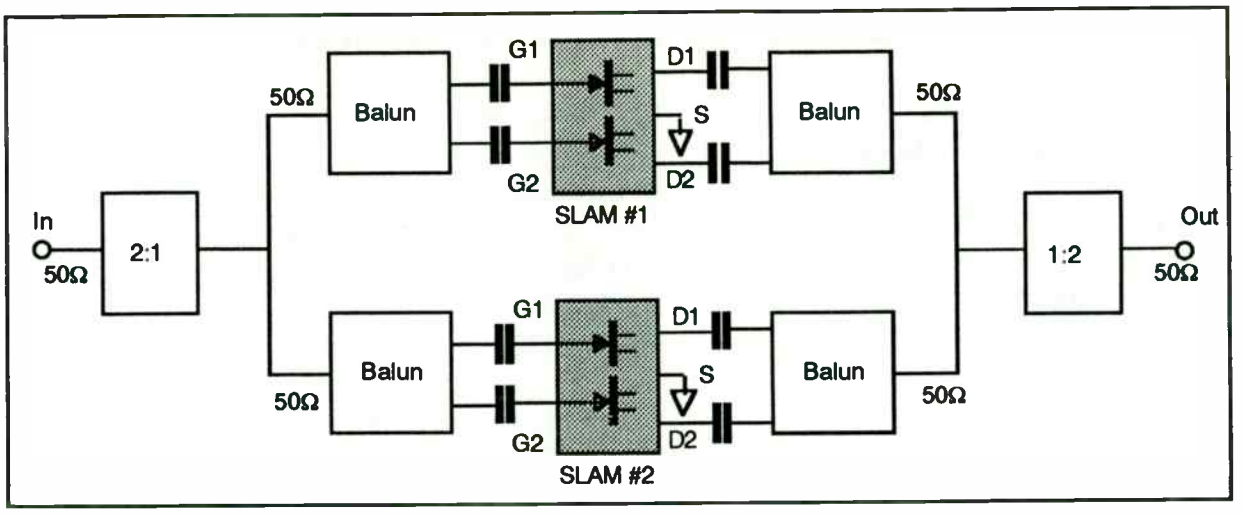


Figure 24. 50W (SLAM0111) or 100W (SLAM0122), 2MHz - 32MHz Amplifier.

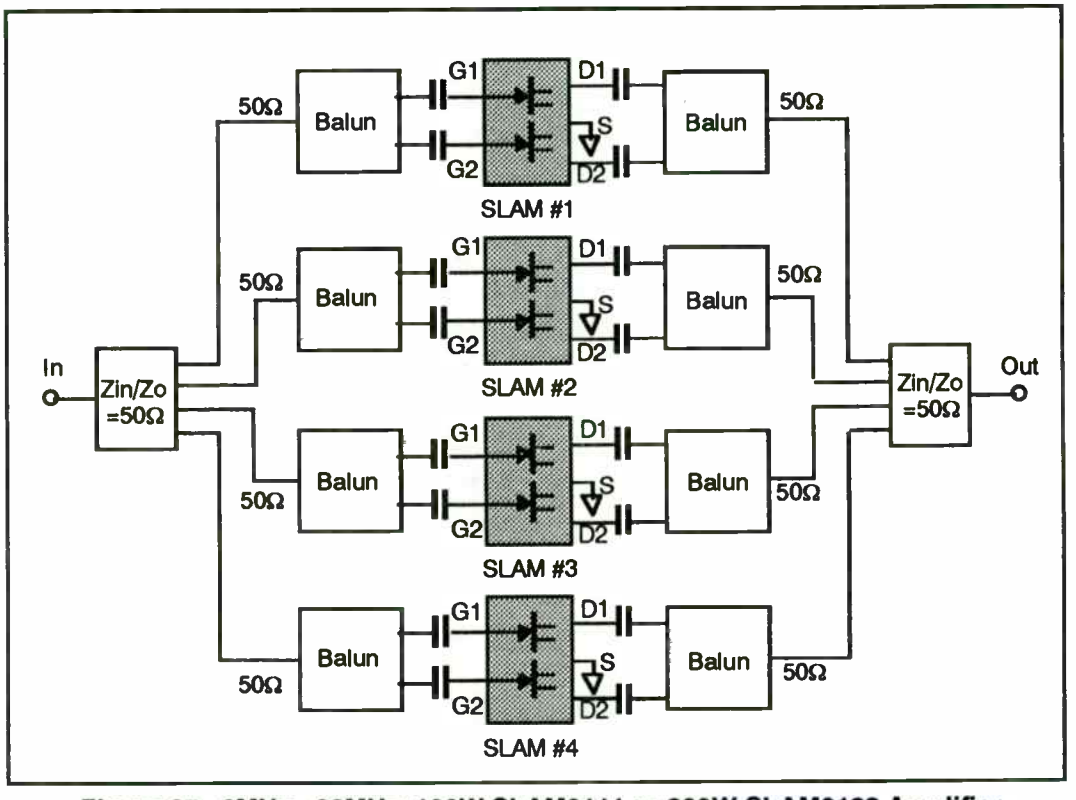


Figure 25. 2MHz - 32MHz, 100W SLAM0111 or 200W SLAM0122 Amplifier.

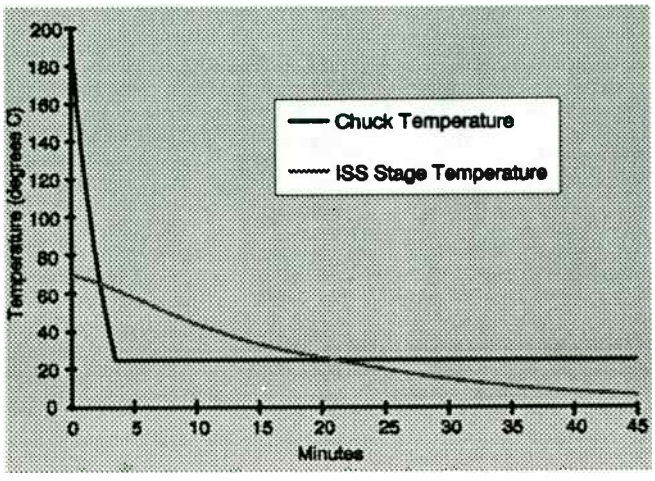
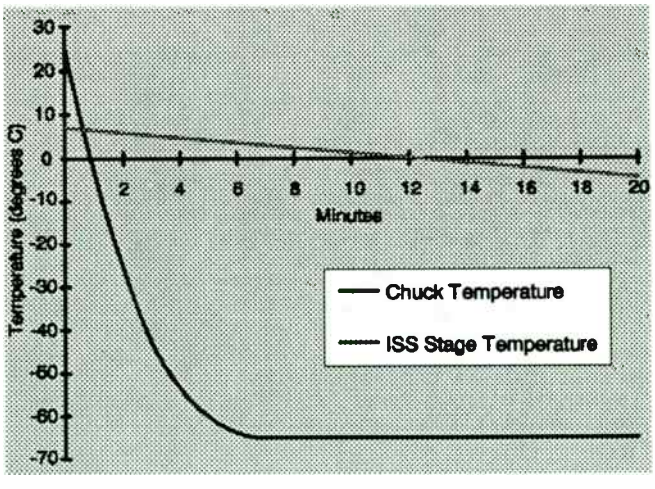


Figure 5. Temperature excursions of wafer stage and ISS stage over time.



Tradeoffs in Practical Design of Class-E High-Efficiency RF Power Amplifiers

Nathan O. Sokal, President
Laszlo Drimusz, Project Engineer

Design Automation, Inc.
4 Tyler Road
Lexington, MA 02173-2404
U.S.A.

TEL. +1 (617) 862-8998
FAX +1 (617) 862-3769
E-mail 73507.247@compuserve.com

Prof. István Novák

Department of Microwave Communications
Technical University of Budapest
P. O. Box 91
H-1502 Budapest
HUNGARY

TEL. and FAX +36 1 181-2968

E-mail t-novak@nov.mht.bme.hu

SUMMARY

A Class E RF power amplifier would have 100% efficiency at any output power and any frequency if it could use *ideal components:* transistors with zero conduction resistance, output capacitance, and turn-on and turn-off switching times; and inductors and capacitors with infinite quality factors (zero parasitic losses).

Real transistors have nonzero conduction resistance, output capacitance, and switching times, and the output capacitance has only finite Q. *Real inductors and capacitors* have parasitic resistance (finite Q) which causes power loss. At a given frequency and output power, the efficiency of an amplifier can be optimized by making a design which deviates slightly from the published nominal design (transistor voltage and time derivative of voltage both zero when the transistor is turned on). In a circuit using real (lossy) components, any deviation from the nominal design will *increase* some components of power loss and will *decrease* others. The efficiency can be optimized by exchanging increases of *some* components of power loss for *larger decreases* of *other* components of power loss, until the *total* power loss (at a specified output power) is minimized. The paper includes equations for all important components of power loss in a nominally tuned Class E circuit. A complete set of analytical expressions does not exist for circuits with *off-nominal tuning* (and probably never will exist), but a circuit simulator can calculate all of the power losses numerically for a specific set of circuit parameters.

The paper explains how efficiency varies with each circuit parameter, and gives transistor figures of merit which quantify transistor power losses in terms of combinations of transistor parameters. With that background established, the paper discusses the practical tradeoffs in optimizing a design which uses nonideal (real) components. For example:

- Reducing the output power reduces transistor-conduction power loss, but can increase output-capacitance discharge power loss, depending on frequency and transistor output capacitance.

- Efficiency can be traded against the flatness of the curve of output power *vs.* frequency across a specified frequency band.

- Output-capacitance discharge power loss can be traded against transistor conduction power loss, at high frequencies where the transistor output capacitance is larger than the capacitance desired at that place in the circuit.

- Increasing the transistor input drive reduces the power loss in the output stage, but increases the dc power consumption of the driver chain (principally the last driver stage). The optimum choice of input drive minimizes the sum of [power dissipation in output stage + power consumption of driver chain].

The optimum tradeoff is specific to each specific design, for its specific set of values of output power, frequency, amplifier bandwidth, and circuit-component loss parameters. For each such specific set of parameters, there is a *best tradeoff* among the different power losses which yields the *lowest total power loss* for a given output power, *i.e.*, the best possible efficiency at that value of output power. For example, if the transistor to be used has *high conduction resistance* and operates at a relatively *low frequency*, the optimum tradeoff will be towards exchanging a *small* increase of capacitance-discharge power loss (proportional to the [*low*] operating frequency) for a *larger* decrease of conduction power loss (proportional to the [*high*] conduction resistance).

Analytical equations do not yet exist (and probably never will exist) for the best-tradeoff off-nominal design, but the best design can be found *numerically* by an optimizer computer program which calls repeatedly on a circuit simulator for answers to "what if?" questions, starting from an initial rough design obtained by using a *qualitative understanding* of the circuit operation and tradeoffs. The HEPA-PLUS computer program yields an optimized design in less than a minute on a 33-MHz 486 IBM-compatible PC. The authors' experience has been that

- performance predicted by the HEPA-PLUS circuit simulator program always agrees well with laboratory measurements, and

- in contests between the HEPA-PLUS optimizer program and the authors' expert design capabilities, *the optimizer program always wins*. One of the authors slaves for two hours on a manual optimization, using the HEPA simulator program to answer "what if?" questions. The resulting "best" design is given to the HEPA-PLUS optimizer program to use as a starting point. In less than a minute, the optimizer program always improves the authors' "best" design by 3 to 6 percentage points.

The paper gives numerical examples of practical tradeoffs, including the predicted performance of an optimized wide-band Class E amplifier which maintains >80% efficiency and ±1 dB variation of output power, across a 1.7:1 frequency band. For that amplifier, the paper shows graphs of output power and efficiency vs. frequency, for the amplifier "before" and "after" optimization. The differences are striking.

For another example of optimizing a Class E amplifier design, including experimental results, see the companion paper at this conference, "Class-E power amplifier delivers 24 W at 27 MHz at 89-92% efficiency, using one transistor costing $0.85," Nathan O. Sokal and Ka-Lon Chu.

BACKGROUND - 1

- Switching-mode RF power amplifiers (Class D and E) can approach 100% efficiency at the design frequency.
- Harmonic content and variations of Pout and efficiency with operating frequency are functions of designer-chosen parameters.
- Power losses are caused by non-ideal parameters of transistors, inductors, and capacitors; can be quantified.

RFW3A020

BACKGROUND - 2

- Design parameters can be chosen to obtain "best" tradeoff.
- Class E retains high efficiency to higher frequencies than does Class D.
- Remainder of this presentation will be about tradeoffs in design of Class E power amplifier.
- Similar tradeoffs can be made in design of Class D amplifier.

RFW3A025

BACKGROUND - 3

Typical Results of Tradeoffs

- Single-frequency design: Get good design from nominal design equations.
  Get about 3-5% higher efficiency from applying tradeoffs.
- Wide-band design, up to 1.7:1 fmax/fmin:
  Flatness of power and efficiency vs. frequency improved by an order of magnitude over a nominal single-frequency design. No design theory yet exists for a wide-band design; automatic optimizer does excellent job.

RFW3A027

APPROACH - 1

- V and I waveforms are easy way to define and understand high-efficiency operating condition. Examine how design choices affect waveforms and performance parameters.
- Explicit mathematical relationships among performance parameters and component parameters for circuit with nominal waveforms.
- Going slightly off-nominal increases some components of P loss, but decreases others. If decrease > increase, circuit efficiency is is higher.

RFE3A030

APPROACH - 2

- Equations show directions of tradeoffs for nominal-tuned circuit. HELPS BUILD UNDERSTANDING of tradeoffs. Extrapolate to off-nominal circuit for qualitative answers.
- For quantitative answers: (a) Tradeoffs change with frequency and with combinations of parasitic-loss parameters. (b) No equations exist for off-nominal operation. Computer analyses or lab experiments are only ways to get quantitative answers.

RFW3040

APPROACH - 3

- Good software tool gives same answers as careful lab tests, but much faster, and gives other information not observable in lab. Aids engineer's understanding.
- Manual method: Identify circuit evaluation parameters; define required performance; vary circuit parameters to achieve required performance with many different sets of circuit parameters; choose "best" set. Better: Use computer method below.
- Define "best" and then find it; use computer.

RFE3A050

PERFORMANCE PARAMETERS

with Defined Ranges of Load,
DC Supply Voltage, & Frequency

- Output power
- Harmonic content
- Variation of Pout across frequency range
- Variation of efficiency across frequency range
- Freedom from parasitic oscillation (many possible causes; sometimes several exist at the same time).

RFE3A060

EVALUATION FACTORS - 1

- Efficiency of entire amplifier chain (Pout/total DC Pin)
- Variation of amplifier efficiency and harmonic content across frequency band.
- Reliability (V, I and P-dissipation stresses on power transistor)
- Product manufacturing cost (e.g., can trade efficiency or stresses for cost).
- Size and weight

RFW3A070

EVALUATION FACTORS - 2

- Design cost and schedule (good design tool helps produce better design and save labor cost and schedule time).
- Evaluate potential designs by weighted sum of evaluation factors.

RFW3A080

BRIEF REVIEW OF CLASS E

- Before discussing tradeoffs, review
  - circuit operating principles
  - high-efficiency waveforms
  - nominal design procedure
- Problem in all switching-mode power circuits: switching power dissipation is proportional to frequency.
- Turn-on dissipation if load is C; turn-off dissipation if load is L; both if load is R.

RFE3A090

SOLUTION ("CLASS E" PRINCIPLE)

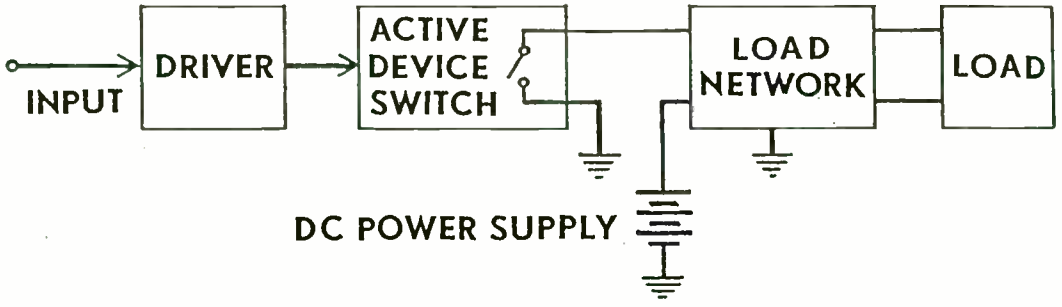
1. Switching power dissipation is i(t)•v(t).
2. Current and voltage each must rise and fall, but they don't have to do it *at the same time*!
3. Time-displace voltage and current transitions; never have high i and high v simultaneously.
4. Don't discharge charged shunt C through switch ([CV**2]f/2).

Result: High efficiency and low stress *even if switching times are considerable fractions of waveform period.*

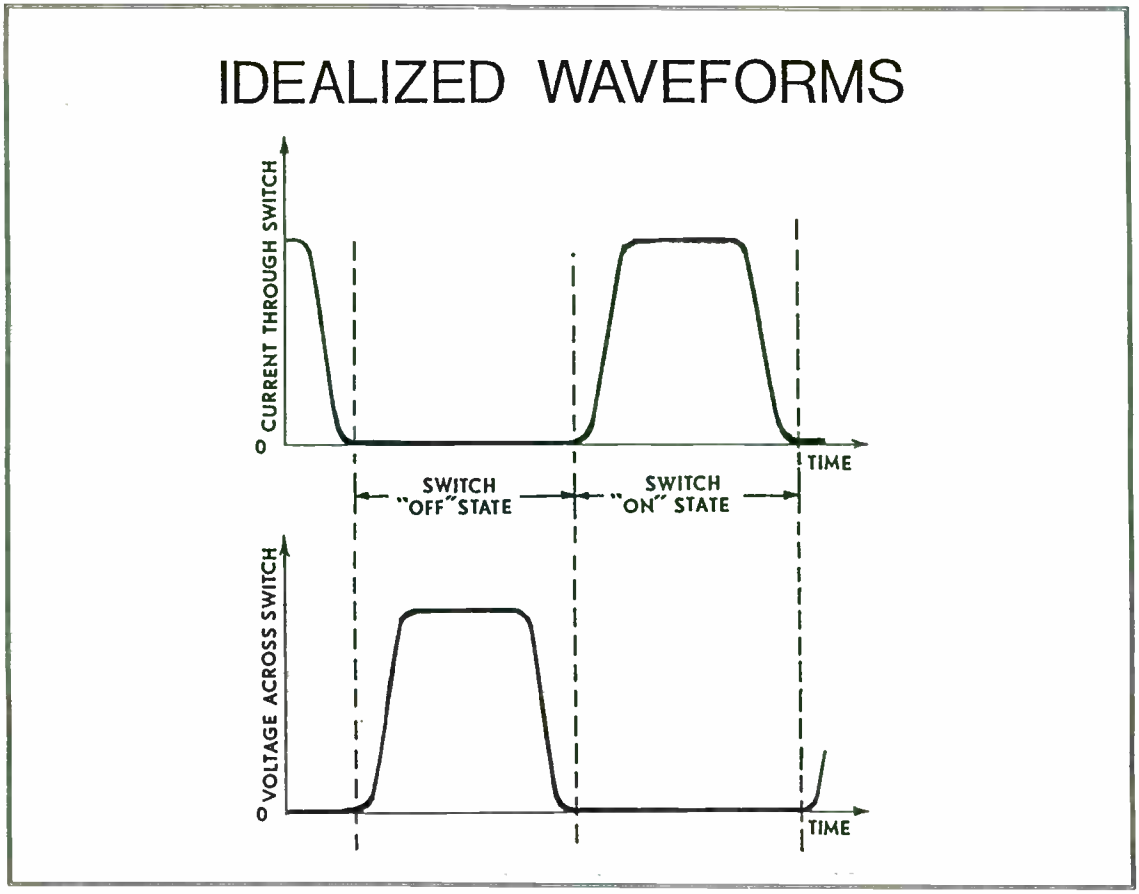
Idealized waveforms on next slide.

RF91W032

POWER-AMPLIFIER BLOCK DIAGRAM



RF91W036



LOAD-NETWORK PROPERTIES - 1

- Excited at input terminals by periodic switch.
- Manufacture time-displaced V and I waveforms at network input terminals.
- Deliver sine-wave current to load at network output terminals.
- Bring network input voltage to zero at switch turn-on time, with zero slope.

RF91W038

LOAD-NETWORK PROPERTIES - 2

- Load network contains any combination of Ls, Cs, transmission lines, magnetic cores, and wire which causes waveform requirements to be met.
- Switching-mode Class F_2 and F_3 are a subset of Class E.
- NEXT SLIDE: Waveforms in published low-order Class E circuit

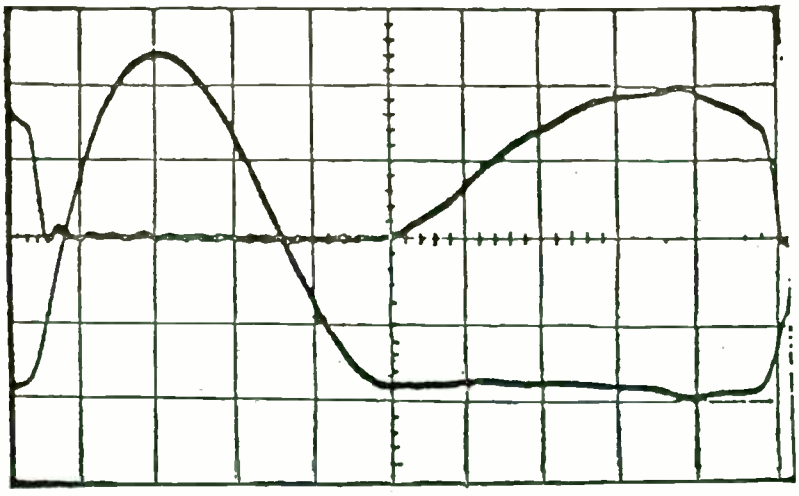
RF91W040

Low-Order Class-E Amplifier

Waveforms

SWITCH CURRENT

SWITCH VOLTAGE

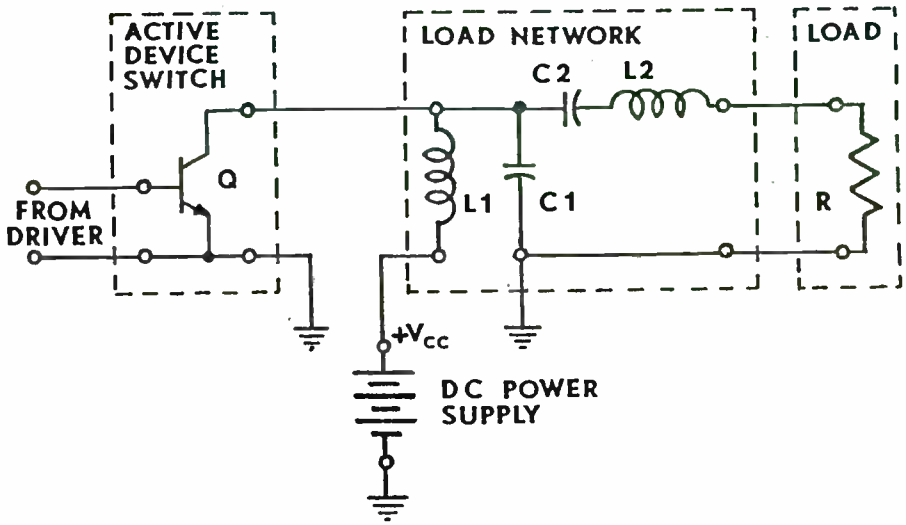


OFF ON

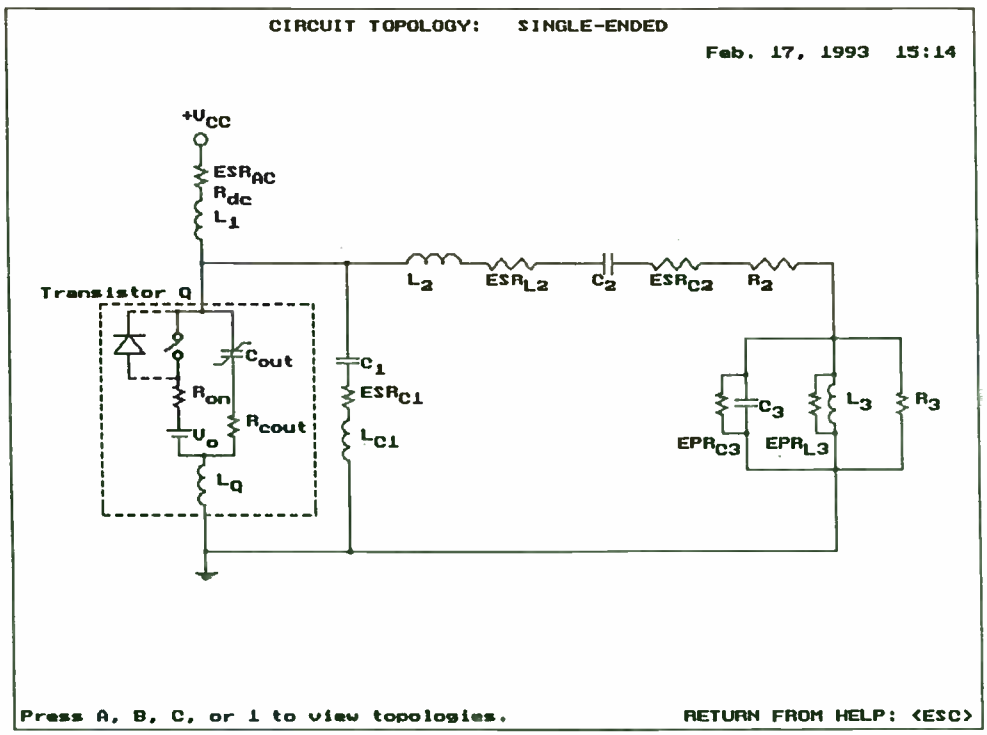
RF1B0200

Low-Order Class-E Amplifier

Schematic



RF1B0190



NOMINAL DESIGN EQUATIONS

1. $V_{CE(pk)} < BV_{CEV}$

2. $V_{CC} = f(V_{CE(pk)})$

3. $R = f(P, V_{CC})$

4. $Q_L \geq 1.7879$, free choice

5. $L2 = Q_L/2\pi f$

6. $C2 = f(R, f, Q_L)$

7. $L1 \geq \ldots$

8. $C1 = f(R, f, Q_L, L1)$

TRADEOFFS - 1

- Efficiency vs. output power for a given transistor and frequency: Ploss/Pout ~ Pout.
- Less Pout increases efficiency, but increases transistor cost/watt output.
- Efficiency vs. frequency for a given Pout: switching losses and C losses ~ frequency (conduction loss does not); first tradeoff multiplier factor increases with frequency.
- Reduce Ron with larger transistor. Improves efficiency but increases cost.

RFW3A120

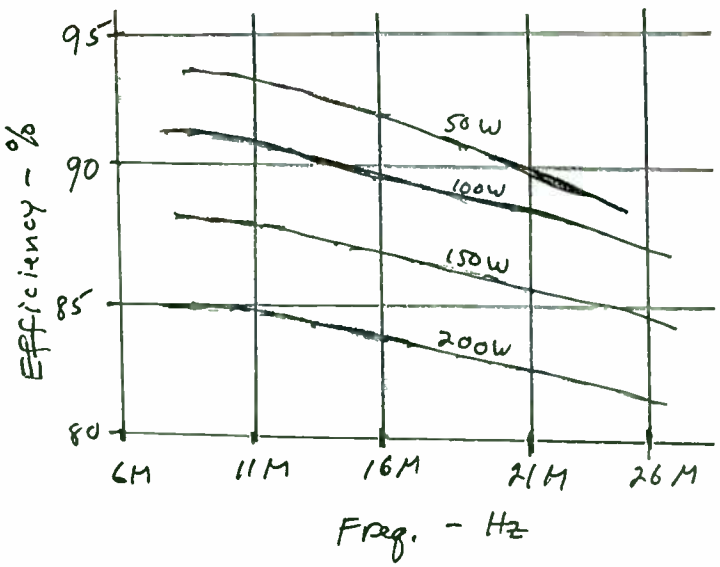
TRADEOFFS - 2

- Limit on larger transistor: When all of C1 is provided by Cout (depends on freq.). Four possible strategies re too-large Cout; sacrifice of eff'y depends on Ron, QL1, and frequency; different combinations --> different "best" choice.
- Result: Tradeoff factors are functions of frequency. "Best" strategy in one frequency region may not be best in another region.
- Results of preceding tradeoffs: graph on next slide.

RFW3A130

TRADEOFFS - 3

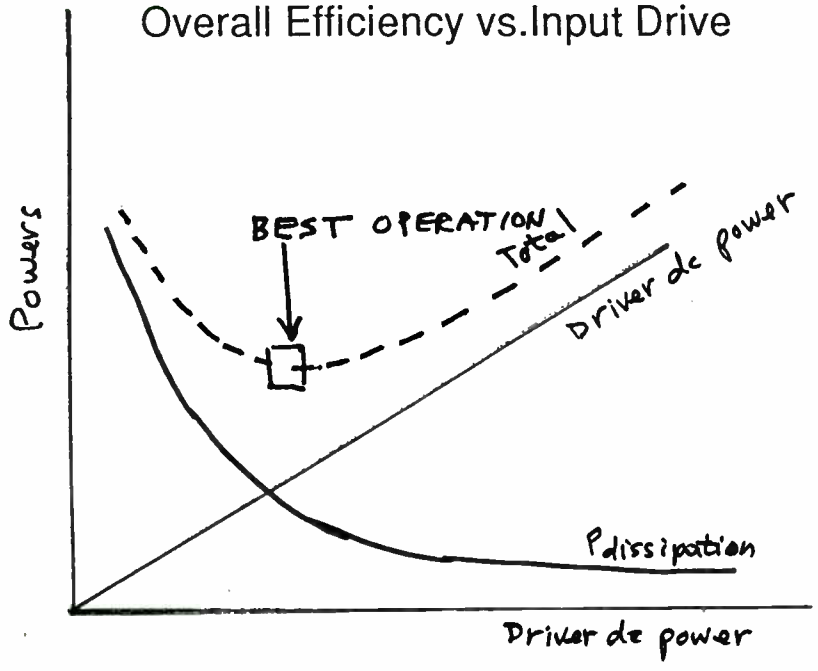
Efficiency vs. Power and Frequency



RFW3A140

TRADEOFFS - 4

Overall Efficiency vs.Input Drive



RFW3A150

TRADEOFFS - 5

- Efficiency vs. deviation from nominal waveforms if Cout > needed at f: i^2(Ron) vs. Cout(dV)^2f/2
- Efficiency vs. required bandwidth = fmax/fmin, harmonic content, and post-filter cost, size, and weight: BW, harmonic content, & needed post-filter ~ 1/QL, but Ploss/Pout ~ QL/Qu of L and C. Power-loss penalty depends on Qu (if high Qu, little penalty for high QL).
- Best wide-band design doesn't have exactly nominal waveforms at ANY frequency in band.

RFW3A160

$\frac{\text{Power loss}}{\text{Power output}}$ with nominal waveforms

---

R_{on}: $\approx 1.37\frac{R_{on}}{R}$

R_{C1}: $\approx 0.2\,\frac{R_{C1}}{R}$

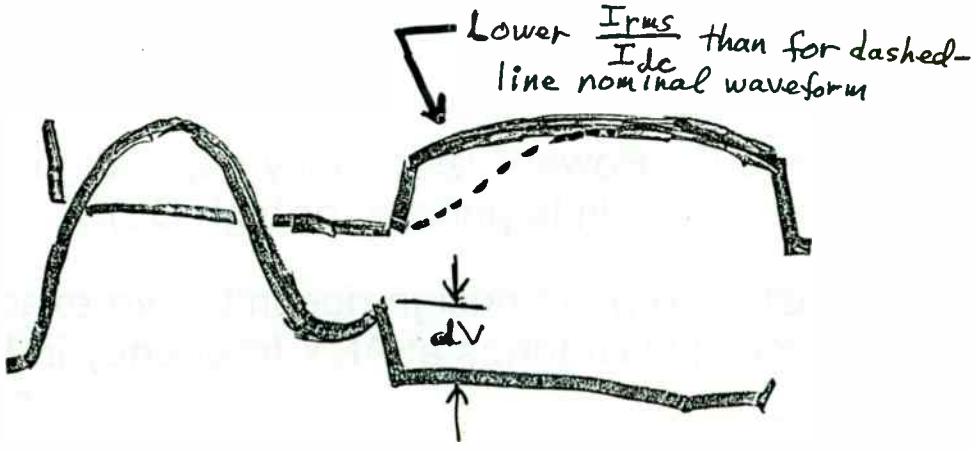
C2 & L2: $\approx Q_L\left(\frac{1}{Q_{UC2}} + \frac{1}{Q_{UL2}}\right)$

Q_u of L1: $a\left(\frac{1}{Q_{uL1}}\right)$

R_{dc} of L1: $R_{dc}\left(\frac{P_{out}}{V_{cc}^2\,\eta^2}\right)$ [η includes effects of all sources of P loss]

RFW3A162

TRADEOFF IF HIGH R_{ON} AND NOT HIGH $\frac{1}{2}C1(dV)^2f$



TRADEOFFS - 6

MOSFET dV/dt CAPABILITY

- Load R < nominal causes drain to swing negative. Then MOSFET substrate diode can conduct.
- High dV/dt during diode turnoff after conduction can cause second breakdown of parasitic NPN (less likely as vendors improve products).
- International Rectifier guarantees 3.5 - 5.5 V/ns capability for HEXFET III products.

RFW3A170

TRADEOFFS - 6 (cont.)

MOSFET dV/dt Capability

- Evaluate capability of your vendor's MOSFETs at your highest frequency and Vcc. If a danger, impose limit on lowest load R.
- This phenomenon does not apply to GaAs MESFETs.
- BJT: next slide.

RFW3A180

TRADEOFFS - 7

BJT Negative Collector

- BJT collector swing negative below "off" base voltage will turn E-B "on" when intended to be "off" --> possible parasitic oscillation.
- Impose limit on low end of range of load R, or connect inverse-diode clamp across C-E.

RFW3A190

BASE OR GATE DRIVER CIRCUIT

- Similar set of tradeoffs for base or gate driver, but influence on overall amplifier performance is smaller by factor of power gain in output stage, e.g., factor of 10.
- Base or gate driver must provide good drive to output stage to ensure good efficiency in output stage. Inadequate drive yields inferior efficiency and can result in parasitic oscillation.

RFE3A200

HOW TO MAKE QUANTITATIVE TRADEOFFS

- Evaluate candidate transistors: HEPA transistor-evaluation module.
- Starting-point design from which to optimize tradeoffs: HEPA preliminary-design module.
- Evaluate the design: HEPA simulation/analysis module (100-1000 times faster than SPICE).
- Automatically optimize the design according to criteria for your specific application: HEPA optimizer (can optimize at up to 16 frequencies in a band, with same or different weighting at each frequency).

RFW3A210

EXAMPLE

- For example of tradeoffs in designing a Class E power amplifier, see paper on 27-MHz PA at this conference. Result: efficiency improved from about 83% to about 90%.
- Computer helped answer many questions about what would be best to do.
- Computer analysis showed that experienced designer's first approach was wrong, and gave a better alternative. Lab tests proved that the computer program was correct.

RFE3A220

Class-E Power Amplifier Delivers 24 W at 27 MHz, at 89-92% Efficiency, Using One Transistor Costing $0.85

Nathan O. Sokal, President and Ka-Lon Chu, Senior Engineer
Design Automation, Inc.
4 Tyler Road
Lexington, MA 02173-2404

Tel. +1 (617) 862-8998
Fax +1 (617) 862-3769
E-mail 73507.247@compuserve.com

SUMMARY

Switching-mode RF power amplifiers (e.g., Class E and voltage-switching Class D) provide significantly higher efficiency than that of Class B and Class C amplifiers. A Class E power amplifier uses a single power transistor, in contrast with the Class D, which uses two or four transistors in a half-bridge or full-bridge topology. (Two 180°-phased Class E circuits can be combined in push-pull if desired, at double the output power of a single circuit.) In all types of switching-mode power amplifiers, (a) the transistors are driven to act as switches at the operating frequency, to minimize the power dissipated while the transistors are conducting current, and (b) the drain or collector efficiency is ideally 100% at low frequencies.

In principle and in practice, the decrease of efficiency with increasing frequency is lower for the Class E amplifier than for the Class D. The penalty for this higher efficiency at high frequency is that the output power per transistor for a given [(peak voltage)*(peak current)] stress is lower for the Class E circuit than for the Class D, by a factor of about 1.5. On the other hand, the Class E circuit avoids the Class-D input-drive difficulties of (a) a large common-mode voltage at the input port of the upper transistor, equal to the full output-voltage swing, and (b) tight tolerance requirements on the relative timing of the switching of the two transistors (with only a single transistor, the Class E circuit has no requirement at all for relative timing between two transistors).

A quasi-complementary half-bridge (two power transistors) voltage-switching Class D power amplifier was reported by F. H. Raab and D. J. Rupp at RF Expo East '92 and in *RF Design*, Sept. 1992. This paper reports on a single-transistor Class E power amplifier.

The Class E amplifier delivered 24 W at 27 MHz, at 89-92% drain efficiency, using a single International Rectifier or Harris Semiconductor IRF520 MOSFET (TO-220 plastic package; price U.S.$0.85 at 100 quantity). This range of efficiency was measured for thirteen amplifiers using transistors from two vendors and three date codes. Circuit simulations established that the best efficiency would be obtained by using one IRF520 transistor, rather than one larger transistor or two IRF520 in parallel.

The presentation includes circuit details and waveform photographs, for the output stage.

SWITCHING-MODE RF POWER AMPLIFIERS
Class D and E

- Class D: pairs of switches, tightly coupled
  Class E: single switch
  Both types: can combine two in push-pull
- Efficiency is ideally 100% at low switching frequencies.
- Efficiency decreases with increasing frequency; less so for Class E.
  Authors' opinion: crossover at a few MHz; specific value depends on the application.

RFW3B010

CLASS D AND E AMPLIFIERS

- Tradeoff for higher efficiency of Class E at high frequency: factor of 1.5 lower value of (Pout per transistor)/(Vpk*Ipk)
- However, Class E avoids Class D input-drive difficulties, namely
  a. large common-mode RF voltage on upper-transistor drive port (the full Vcc)
  b. tight tolerance requirement on relative timing of switching of the two transistors.
  These difficulties become more formidable with increasing frequency. Quit about 10 MHz.

RFW3B020

CHARACTERISTICS OF CLASS E

- Transistor operates in switching mode; each transition time up to 15% of the RF period.
- Switch turn-off voltage is low, giving low loss during slow turn-off.
- Switch turn-on current is low, giving low loss during slow turn-on.
- Allowing slow turn-on and turn-off makes Class E work well at high frequencies, up to 15% of 1/(turn-off transition time).

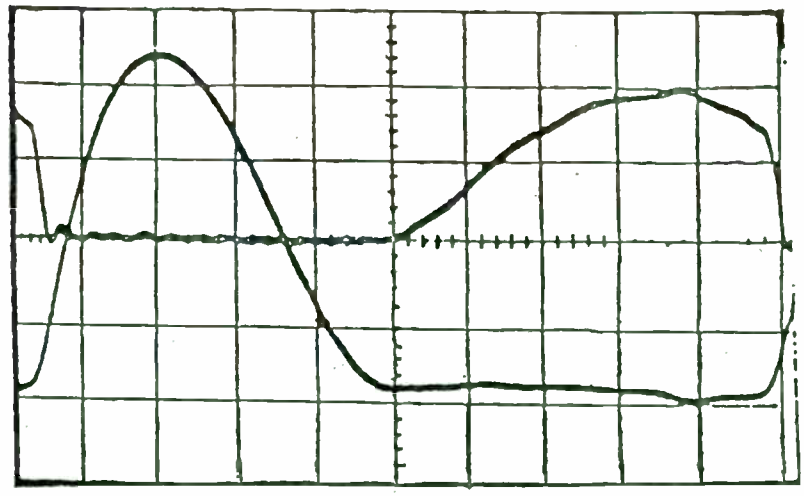
RFW3B040

Low-Order Class-E Amplifier

Waveforms

SWITCH CURRENT

SWITCH VOLTAGE

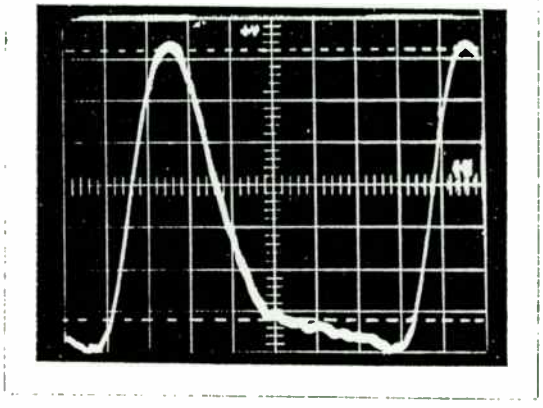


OFF ON

RF1B0200

27-MHZ CLASS E AMPLIFIER

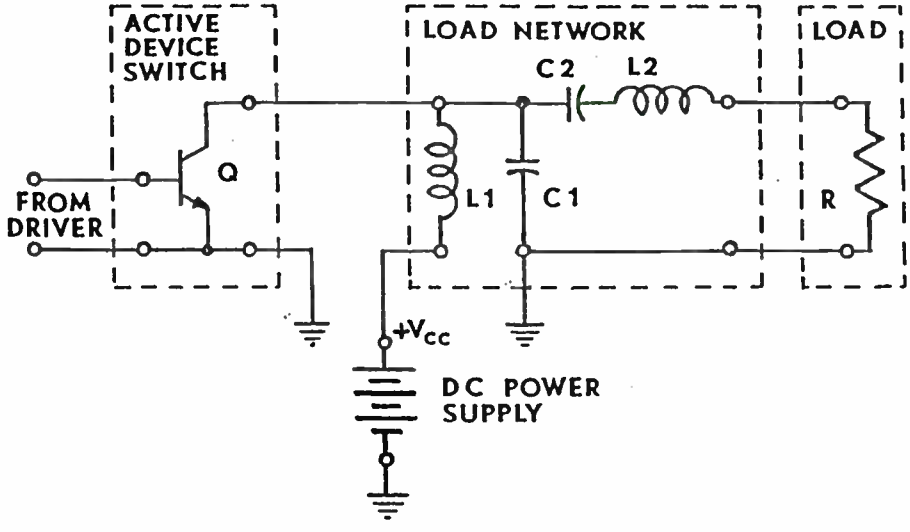
Drain-voltage waveform



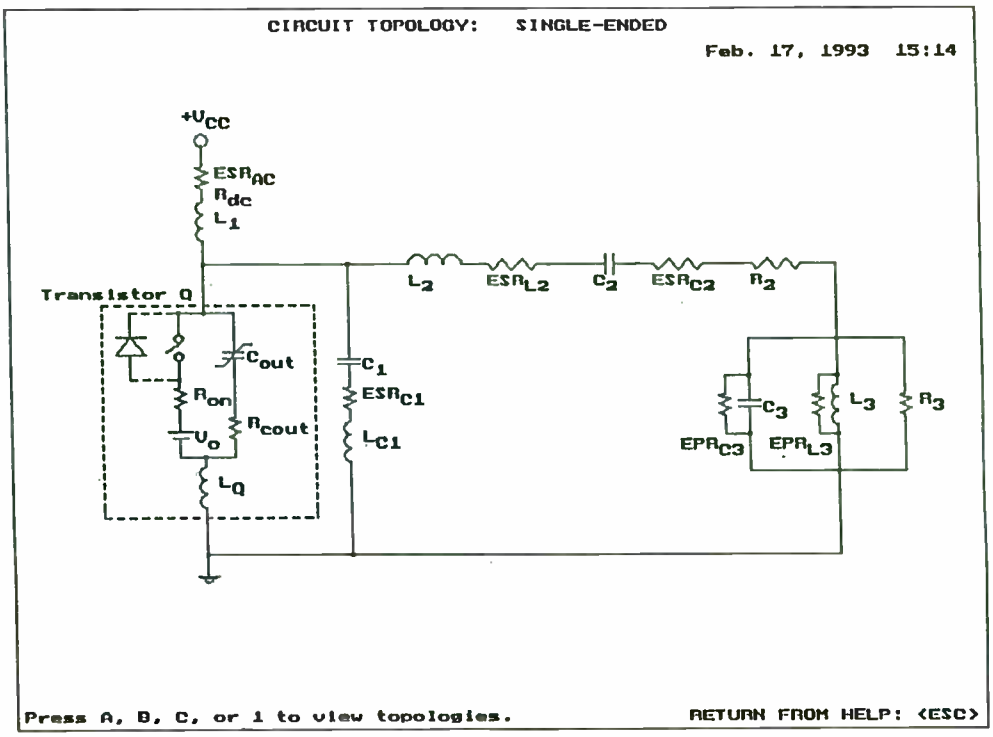
RFW3B055

Low-Order Class-E Amplifier

Schematic



RF1B0190



RF-GENERATOR REQUIREMENTS

(Master oscillator/power amplifier)

- Frequency: 27.12 MHz
- Output power: 20 W minimum with 24 Vdc from battery pack; change power by changing taps on battery pack.
- Overall efficiency: >80%
- Low product cost: use low-cost plastic-packaged transistor.

RFW3B070

TRANSISTOR CHOICE - 1

- MOSFET instead of BJT
  a. effects of unspecified parasitic parameters are less important
  b. potential efficiency is higher
  c. easier to drive input
  d. easier to understand (fewer arcane effects)
  e. less problem if accidental overstress
- Required voltage rating: 3.5 x 24 Vdc x safety factor = 100 V
- Low-cost, in plastic package: International Rectifier IRF510 series (multiple sources)

RFW3B080

TRANSISTOR CHOICE - 2

Which 100-V transistor?

- Two IRF510 or 520 in parallel to halve the parasitic lead inductances (expect problem)
- IRF530
- IRF540 is too large (suitable for 200 W).
- Compare computer-predicted results (HEPA program) and choose best tradeoff.
- Tradeoff is lower Ron (less conduction loss) vs. higher Cout (more turn-on loss).

RFW3B090

TRANSISTOR CHOICE - 3

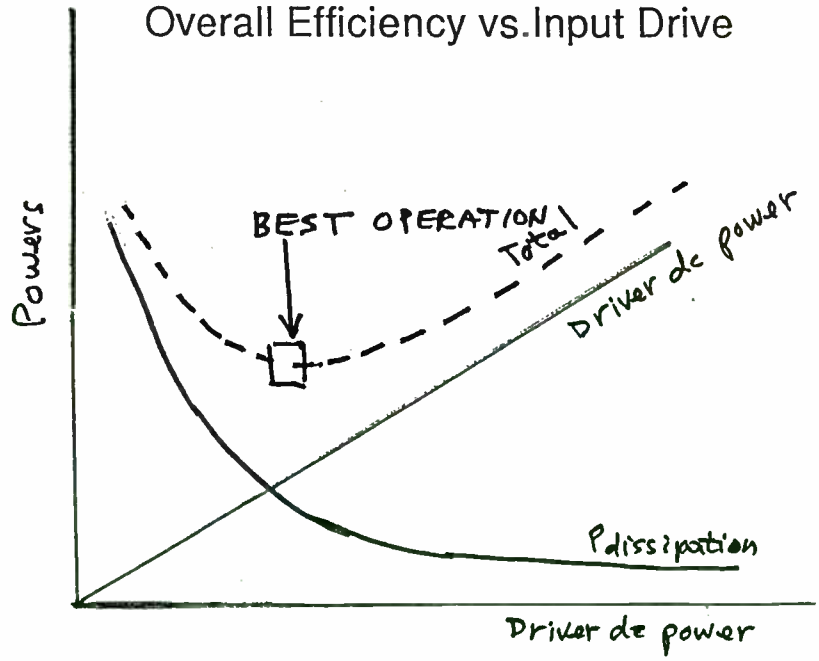
Comparison Results

- Computer optimization gives best efficiency with two IRF520 in parallel.
- Experimental results: Effects of parasitic inductance much less than feared, due to careful low-L PC layout.
- But driver stage has heavy load of doubled inputs of output stage; uses more driver P.
- Try single IRF520 to trade slightly higher loss in output stage for less driver power.

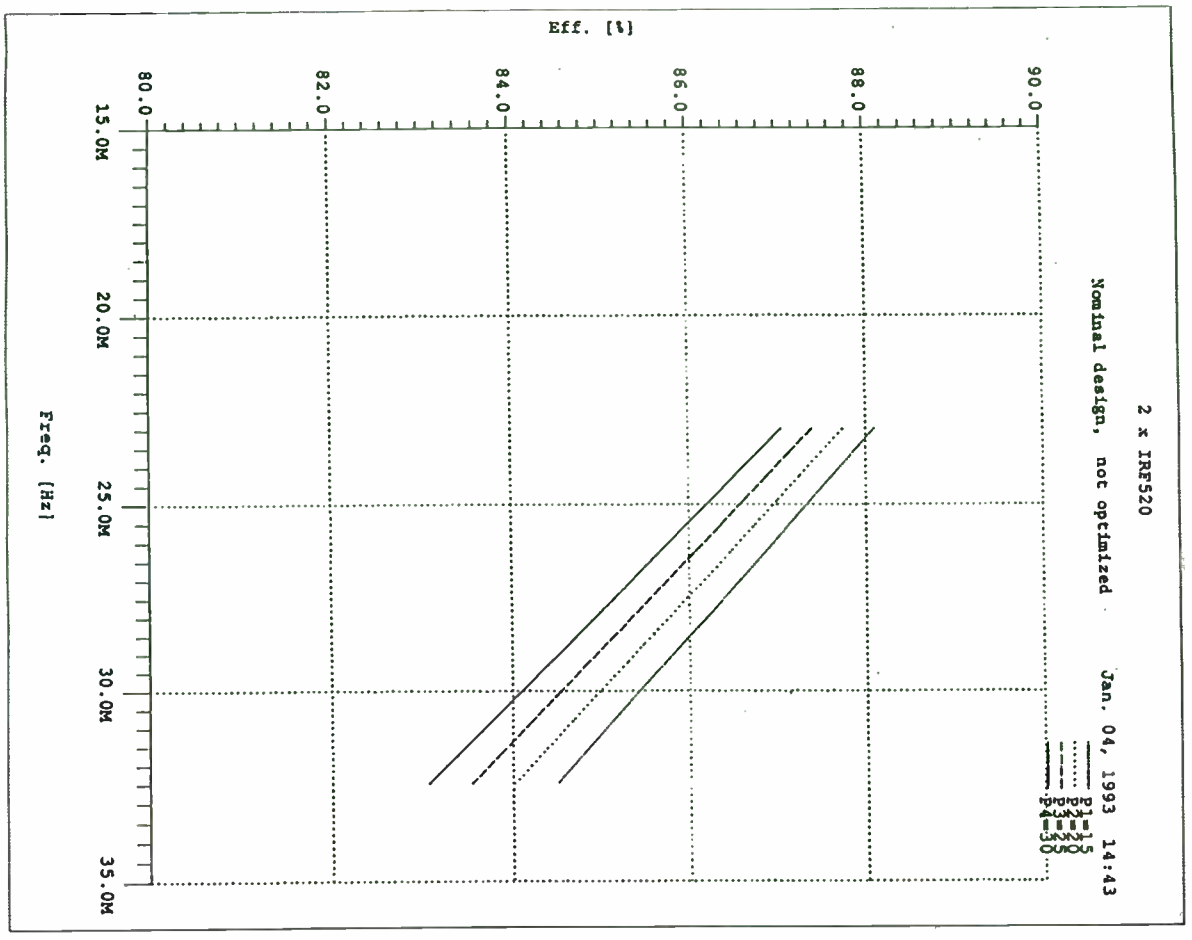
RFW3B100

TRADEOFFS - 4

Overall Efficiency vs.Input Drive



RFW3A150
B102



High-Efficiency Power Amplifier PLUS (HEPA II+/WB) Jan. 04, 1993 16:51

1 x IRF520

ENTER CIRCUIT PARAMETERS and/or TITLE

Common-Source TRANSISTOR with diode		LOAD without filter	
Frequency (f).....[Hz]:	2.712E+07	Load location, R3 or R2?	R3
Duty ratio (D).....:	0.5	Load resistance..[ohms]:	50
"On" resistance...[ohms]:	0.25	R3/Rload.....:	1
Vo[V]:0	Lo [H]: 6E-09	L3.....[henries]:	1E-05
Transition times: turn-on,	turn-off [s]	Qu:10000 at freq:	2.694E+07
on:3.5E-09	off: 3.5E-09	C3.....[farads]:	2.3509E-10
Cout.....[farads]:	1.3E-10	Qu:200 at freq:	2.694E+07
Cout series resis.[ohms]:	0.625		

LOAD NETWORK, Single-ended

DC supply (Vcc)..[volts]:	24.	C2.....[farads]:	2.5E-09
L1.....[henries]:	3E-06	Qu:200 at freq:	2.694E+07
Qu: 50 at freq:	2.694E+07		
Rdc.....[ohms]:	0.051469	Network loaded 0 or L2?	L2
C1.....[farads]:	1.4683E-11	L2.....[henries]:	2.053E-07
Qu: 200 at freq:	2.694E+07	Qu:200 at freq:	2.694E+07
Lc1.....[henries]:	0	Network loaded Q.....:	3.2176

,) ENTRY: ALPHANUM. COMPUTE: (PgDn) MAIN MENU: (ESC)

SELECTING PARAMETER...

High-Efficiency Power Amplifier PLUS (HEPA II+/WB) Jan. 04, 1993 16:51

1 x IRF520

EFFICIENCY, POWERS, and STRESSES

Single-ended

Collector/drain efficiency.....(Pout/Pin)		89.897%
Collector/drain inefficiency...[(Pin-Pout)/Pin]		10.103%
Overall efficiency.....(Pout/(Pin+Pid)]		82.101%
DC power input (Pin).....(watts]		26.326
Input-drive power (Pid).....(watts]		2.5
Power output (Pout).....(watts]		23.667
Power loss in L1.....(watts]		0.088116
Power loss in L2.....(watts]		0.41123
Power loss in C2.....(watts]		0.027872
Resistive power loss of transistor & C1 (watts]		1.1856
Turn-off power loss of transistor.....(watts]		0.64045
Turn-on power loss of transistor.....(watts]		0.070132
Power loss in L3.....(watts]		6.8523E-05
Power loss in C3.....(watts]		0.23622
Output voltage, current (at Rload) [V,A]:	34.4	0.68799
Transistor peak voltages....(volts]: normal	79.559	inverse None
Transistor peak currents..(amperes]: normal	6.5714	inverse None

'DISPLAY RESULTS' MENU: (ESC)

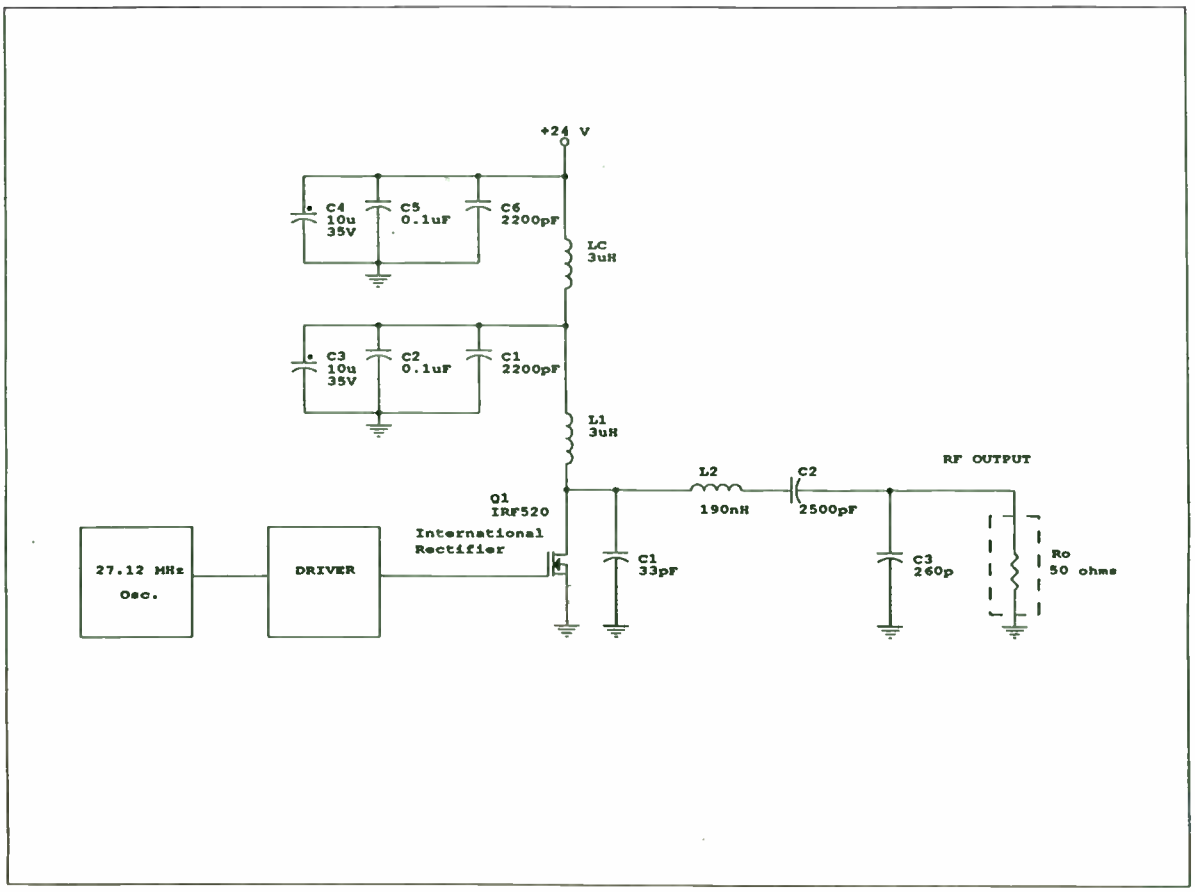
DISPLAYING COMPUTED POWERS...

TRANSISTOR CHOICE - 4

Experimental Results

- Output stage drain efficiencies of one or two IRF520 were as predicted by HEPA program.
- Final design: one IRF520 for 24 W RF output.
- Single IRF520 had slightly lower drain efficiency, but power saved in driver stage more than made up for the slightly higher drain power loss.
- Fifteen assemblies with IR and Harris transistors had drain efficiencies of 89 to 92%.

RFW3B110



EXPERIMENTAL RESULTS

Using One IRF520, 24 W, 27.12 MHz

- Measured drain efficiency = 89 to 92%. (HEPA computer program predicted 90%.)
- Overall efficiency = 84%, including dc power to crystal oscillator and driver chain.

RFW3B120

A UHF/L-Band FET Module for Pulsed Power Avionics Applications.

Frank Sulak, Ken Sooknanan,
Toru Nakamura, Al Rosenzweig,
Adrian I. Cogan

MicroWave Technology, Inc.
4268 Solar Way
Fremont, CA 94538

Summary

A large number of applications exist for pulsed solid-state power in the frequency range between 800 MHz and 1300 MHz. These include modern avionics systems, such as distance measurement equipment (DME), air-traffic control transponders, identification friend or foe (IFF), and secure communications systems such as the Joint Tactical Information and Distribution System (JTIDS). These systems, as well as air traffic control radars, are being designed to communicate more and more information in digital format or with sophisticated pulse trains requiring that the power amplifiers reproduce the modulation accurately while occupying less volume and using less DC power.

This paper describes a new technology used in the design and fabrication of very small size pulsed operation, high power, broadband and narrow band amplifier modules for avionics applications. These Compact Amplifier Modules (CAMs) exhibit high peak power (50 W through 250 W) in a very small size and operate broadband, from 800 MHz through 1250 MHz.

The CAMs use power silicon FETs called Solid State Triodes™ (SSTs). The main design aspects related to such high power pulse SST circuits operating at L-band frequencies are discussed, and practical CAM fabrication issues are addressed.

In high pulse power avionics applications, the SST CAMs have unique performance advantages over bipolar transistors. CAMs exhibit wide dynamic range of gain control, wide dynamic range for linear gain, excellent thermal stability, and have very short pulse rise and fall times.

CAM applications and performance figures are presented and discussed in the last section of the paper. While the SST CAMs were developed for avionics applications, the technology can be used for other areas where tens and hundreds of watts of pulsed RF power are required over narrow band or broadband frequencies ranging from 500 MHz through 1300 MHz.

Introduction: The SST.

In recent years, power Bipolar Junction Transistors (BJTs) have seen their virtual monopoly in RF applications challenged first by MOSFETs and recently, by Solid State Triodes (SSTs). While at RF and VHF frequencies MOSFET use keeps increasing at the expense of BJTs, the BJT domination at UHF frequencies and into L-band has continued. The SST, has now demonstrated performance levels superior to BJTs and MOSFETs.

SSTs are depletion-mode silicon junction FETs which exhibit unsaturated, triode-like I-V characteristics, similar to those of a vacuum

triode (Figure 1). The SST requires bias conditions similar to a vacuum tube or a GaAs MESFET: negative gate voltage and positive drain voltage.

As with any other RF transistor, the SST electrical performance, and parameters such as power gain, efficiency, output power, etc. , are dependent on the operating point.

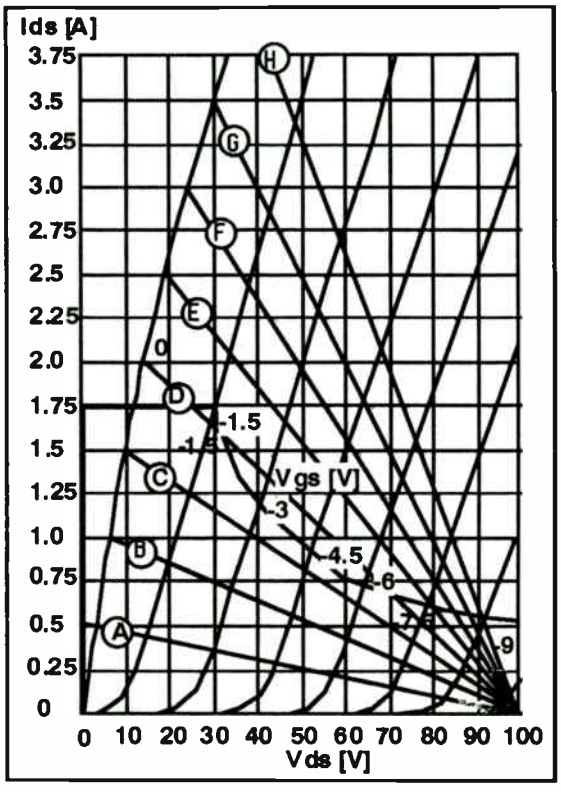


Figure 1. SST I-V Characteristics.

The maximum power available from a given transistor active area remains constant and is limited by thermal dissipation and is bounded by a maximum (breakdown) voltage and current. The resulting safe operating area (SOA) widens as one moves from CW into pulse operating conditions. The shorter the pulse and the lower the pulse duty factor, the wider the device SOA. Ultimately, a maximum allowed current density and the junction breakdown voltage define a rectangular SOA for BJTs and MOSFETs.

For the BJT, exceeding the maximum specified current will result in a significant decrease in current gain, while for the MOSFET, attempts to further increase the drain current by increasing the gate voltage will result in gain compression and, if Vgs is further increased, the gate oxide will rupture.

Unlike the BJT and the MOSFET, the SST pulsed SOA extends beyond the shaded CW SOA boundary in Figure 1. As a consequence, in selecting the load impedance for maximum output power, one can access a wider region of the I-V characteristics. As shown in this figure, the SST transconductance keeps increasing with the operating current. One consequence of this unusual operating behavior is that higher pulsed power levels can be achieved from the same silicon active area when comparing SSTs to BJTs or to MOSFETs. The pulsed power to CW power ratios of better than 6 dB are routinely measured with present SSTs.

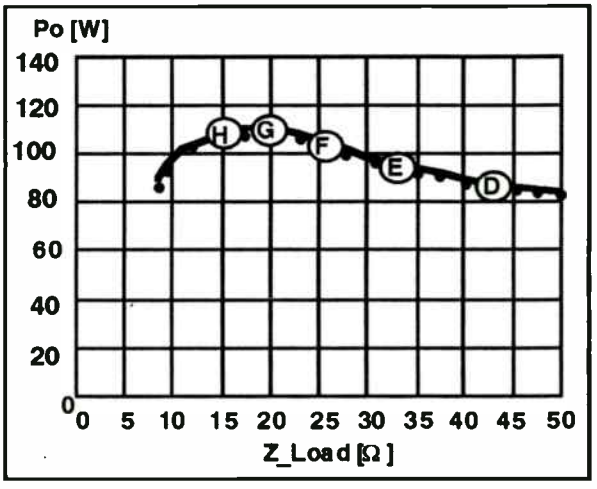


Figure 2. Pout vs. Load Impedance.

Several load impedance values are marked in Figures 1 and 2. The calculated dependence of the maximum available output power on the load impedance is plotted in Figure 2. This illustrates the relatively slow Po vs. Z_L dependence. This is another SST advantage since in the case of BJTs and MOSFETs the maximum power load impedance is always selected to run from the 'knee' of the collector or drain current to the off-current maximum voltage point.

MwT has performed extensive SST testing under RF pulse operating conditions at frequencies ranging from 500 MHz through 1300 MHz. It should be noted that, to date, no power MOSFETs compete with BJTs and

SSTs for this type of applications. Some of the most relevant SST advantages for pulsed applications are as follows:

- The SST can switch RF pulses with rise and fall times of less than 10 ns. By comparison, BJTs usually switch in tens of nanoseconds.

- The SST excellent linearity allows for the RF pulse spectral response to be better preserved or to be easier shaped to desired specifications. In present BJT applications, the pulse spectral response is tailored by modulating the transistor collector voltage at high current levels, a very difficult task to implement. By contrast, in a SST CAM power control requirements are easily met in a low current, high impedance environment.

- The SST has a wide dynamic range of linear gain control, typically more than 10 dB. This feature provides previously unachievable control levels in avionics high power RF applications presently relying on PIN diode or mechanical switches. On-Off RF power ratios of better than 40 dB can easily be achieved.

- The SST is thermally stable as the drain current will decrease slightly with an increase in temperature. It also has no power drift after turn-on and its power derating coefficient is approximately 0.01 dB/°C. One consequence is that simpler thermal compensation circuitry is required. SSTs operation was tested at junction temperatures ranging from liquid helium (3°K) through 300°C. Both the BJT and the MOSFETs cease to operate at such low temperatures.

- The SSTs are extremely rugged transistors, withstanding VSWR figures in excess of 20:1. This ruggedness is due mainly to the SST JFET structure, which has fewer parasitic components when compared to BJTs and MOSFETs and does not have very thin junctions or thin gate oxides, which are easily degraded by over-voltage, over-current, or high operating temperatures.

- The SST is its inherently radiation hard. Due to the simplicity of its structure, when irradiated, the SST does not exhibit a current gain decrease such as with the BJT, or threshold voltage shifts, such as with the MOSFET. SSTs are superior in both total dose and radiation bursts environments. While this is not directly an electrical circuit advantage, it will promote SST use in applications which require the electronics to operate under harsh ionizing radiation conditions.

Compact RF Power Amplifier Modules.

Conventional high power pulsed RF amplifiers use discrete transistor packages mounted on external heat sinks, with softboard printed circuits, and mostly lumped discrete passive components. While this is a convenient and well proven approach, it does not make maximum use of the transistor capabilities. These amplifiers are also relatively bulky.

Microwave Technology (MwT) has developed a hybrid circuit technology suitable for the fabrication of high power, compact RF amplifier modules. Based on its present thin film hybrid IC technology, the new circuit modules eliminate the need for discrete RF power packages and present the user with a very convenient Zin = Zout = 50 Ω solution.

These modules have the internal matching elements fabricated in the immediate vicinity of the SST chip. Near chip matching increases circuit efficiency by reducing loss associated with power combining and broad-banding and also improves system reliability. Costs are usually reduced, when compared to the present single device unit cost and circuit fabrication costs for discrete L-band power transistor amplifiers.

The first step in implementing these compact SST power amplifiers is the Aluminum Nitride Carrier Assembly (ANCA). An ANCA consists of a high thermal conductivity rectangular ceramic pill with metallized "via" grounding holes, on top of which thin film distributed

impedance matching element circuitry is deposited and patterned.

A single-ended ANCA and its equivalent circuit are shown in Figure 3.a. As shown, the first output impedance matching step uses a shunt inductor decoupled through a low reactance capacitor. Low-pass LC sections are used on the SST input side and are easily recognized in the figure. As shown, the first matching elements are placed in close proximity to the SST chip, thus reducing parasitic component contribution and the loss due to the high circulating currents in the shunt L.

Additional input and output matching sections are used to achieve the desired output power over the specified bandwidth. These distributed element circuits are fabricated on low loss alumina, also using thin film processing technology.

circuits and RF bypassing elements inside a small size housing, one ends up with a Compact Amplifier Module (CAM), such as the one illustrated in Figure 4. These less than 2" long CAMs deliver 150 W of peak power at 1000 MHz with over 30 % bandwidths.

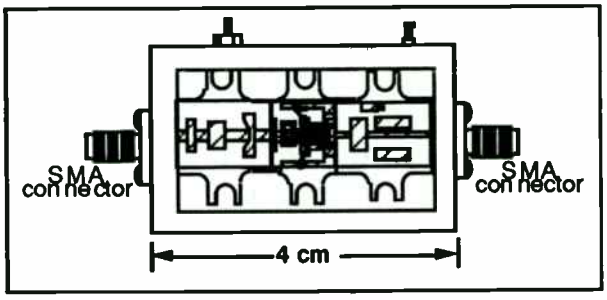
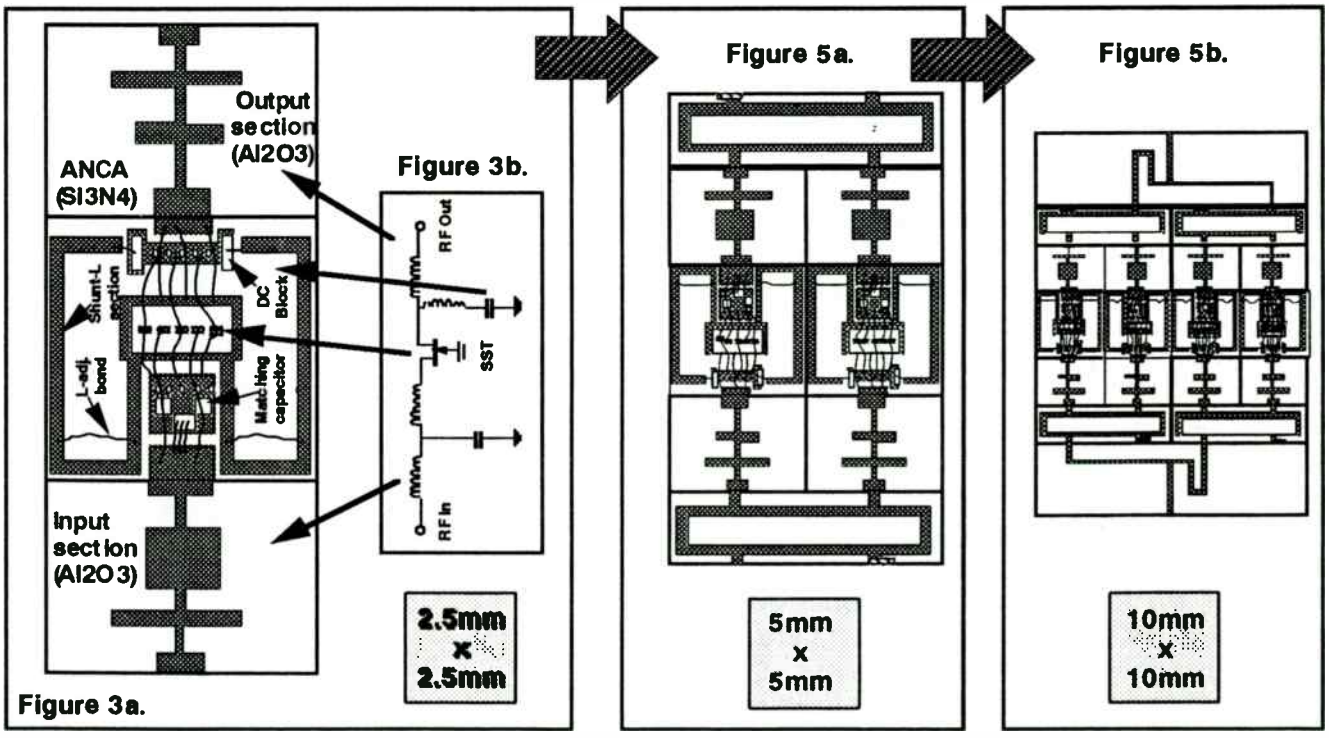


Figure 4. 100W CAM.

Two ANCAs and their matching circuits are power combined into a higher power CAM using Wilkinson or 90° hybrid couplers, as illustrated in Figure 5.a. Further combining will yield Power Amplifier Modules (PAMs) capable to deliver 300 W - 500 W of peak RF



CAM Components: ANCA +Iinput and Output Mat ching Sections

200W CAM Layout.

400W PAM Layout.

When an ANCA and its corresponding input and output matching circuits are assembled together with the appropriate DC biasing

power (Figure 5.b.). The PAM operating frequency can be centered at any frequency

in the 800 MHz through 1000 MHz with a bandwidth that can exceed 30%.

The CAM design, fabrication and characterization work proceeds by addressing one matching section at a time, and moving from the SST die plane toward the module outputs, and begins with the output shunt-L section.

In order to facilitate the design of the ANCA and CAM matching circuits, RF characterization proceeded in two steps: a) before connecting the matching elements (active SST die deembeding) and b) with shunt matching elements connected. As shown in Figure 3., bonding wires are used to tune the ANCA section of the matching circuit.

ANCA and CAM device characterization was carried out over the 500 MHz - 1300 MHz frequency range. Relying on test data acquired during the above evaluation steps, single and then two stage input matching elements were added and their contribution characterized. 'On board' DC feed circuit elements were also included, and designed to provide the best pulsed mode operation characteristics.

Presently, CAMs are fabricated with the ANCAs and the matching circuits mounted inside nickel plated aluminum or OFHC copper housings, which are hermetically sealed. SMA and DC feedthrough connectors allow for RF, DC power and control signals to be fed to the amplifier (Figure 4).

SST Biasing Considerations.

As mentioned, the SST is a depletion mode JFET. When the source is DC ground and the drain is positively biased, the device requires negative gate voltage to be turned off. In order to preserve the SST performance into the 1400 MHz frequency range, the device has to be operated common-gate. Common-gate, the SST exhibits an extremely low feedback capacitance (Cds), of approximately 0.1 pF/W, assuring a virtual separation between the SST input and output sections.

Figure 6 shows the SST biasing circuit used for pulsed RF operation. The circuit provides a DC pulse input and an independent adjustment of the device quiescent current. Between the RF pulses, the SST is turned off by adjusting the gate voltage to the appropriate value of V_{GS}. During the RF pulse, the NPN transistor is turned on, thus setting the gate voltage to a less negative value with respect to the source.

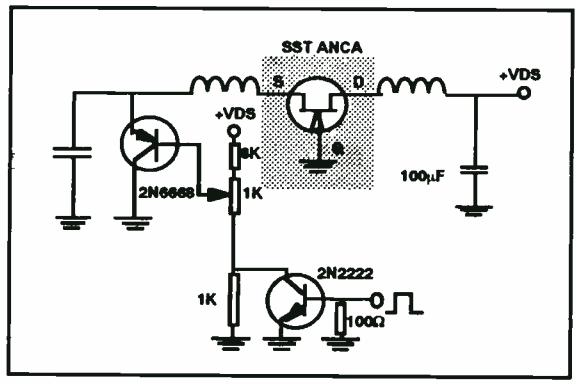


Figure 6. CAM Internal Biasing Circuit.

This circuit allows one to set the SST operating point for class A, B or C operation for the duration of the RF pulse. The same circuit can be used for CW RF operation if the NPN transistor is eliminated or kept on.

Pulsed Power Avionics Applications.

Some of the most commonly known applications in avionics equipment utilize the 800 to 1400 MHz frequency band and fall into the following categories:

- 960 to 1220 MHz - DME & JTIDS
- 1030 & 1090 MHz - IFF
- 1200 to 1400 MHz - Radar

These systems operate in a pulse modulated RF signal mode and information is carried in the pulse characteristics. Distance information is derived from signal travel time while additional information resides in the pulse coding.

DME and IFF systems basically have been operating in narrow pulse, low duty cycle modes (<10 µs and < 10%). JTIDS requires

high duty capability for some given time intervals while some of the radar systems operate with fairly long pulses.

Linear pulse envelope modulation characteristics are of primary interest for DME operation while other applications require fast rise and fall time capability. The compact amplifier modules can satisfy all of these requirements.

Experimental Performance Figures.

Most of the MwT Pulsed RF CAM work is done at the JTIDS, DME, and IFF frequency ranges. Typical experimental results are presented below. The module RF performance is evaluated using the computer controlled test setup of Figure 7. RF output power , power gain and efficiency are determined over various pulse and biasing conditions and over a wide temperature.

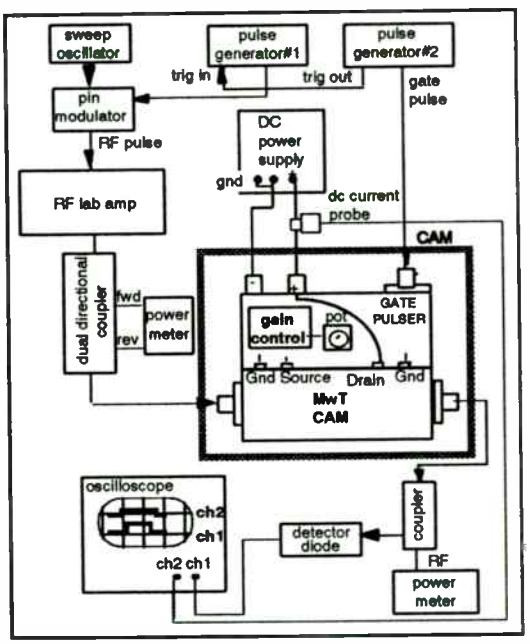


Figure 7. CAM Test Setup.

The equipment shown is linked by the GPIB bus and controlled by an IBM compatible PC. The system is capable of both pulsed and CW measurements. Pin vs. Pout at any frequency in the 500 MHz to 1300 MHz band can be measured to obtain linear gain, one-dB gain compression, and saturated power output. CAM frequency response vs. gain, Pout, return loss, and efficiency can be quickly measured on this automated test station. All data generated can be saved in non-volatile memory for future use.

Calibration of the system is also automated. The coupling factors of the input dual-directional coupler, any insertion loss of the connectors, and the attenuation of the load are all measured vs. frequency and stored in a calibration file. This data is then used to correct the raw measurement data.

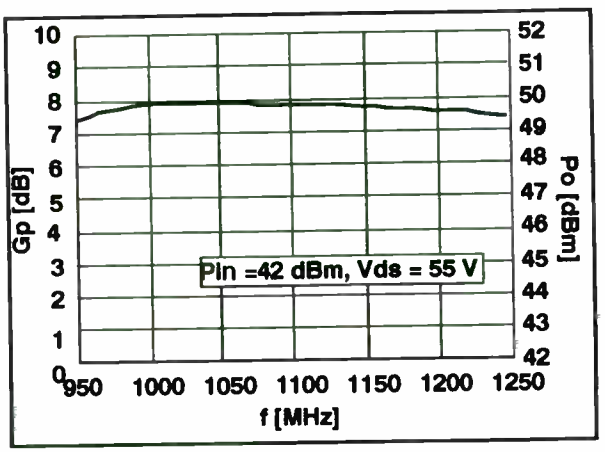


Figure 8. 100 W CAM: Gp vs. frequency.

Figure 8 shows the frequency dependence of an JTIDS CAM rated for 85 W. This CAM has a power gain of 7.5 dB ± 0.5 dB across the 960 MHz - 1220 MHz frequency band. The output power dependence versus the input power (Figure 9) for the same CAM illustrates the wide linear power dynamic range of approximately 50 dB.

Data taken on a 65 W broadband CAM, tested over the 800 MHz through 1100 MHz frequency range (Figure 10), demonstrates a module efficiency of 40 % across the whole bandwitdh. As shown, this CAM has a nearly flat power gain of 8 dB.

Typical CAM power gain vs. output power data for a 50 W CAM is plotted in Figure 11. The figure also contains input return loss data. As shown, when the input power increases and the SST moves from class A and into class AB and than B operation, the device power dissipation is reduced as efficiency increases. This change, combined with the fact that the module was designed for best

match at the higher power levels, induces a small increase in the voltage gain as the CAM approaches power saturation.

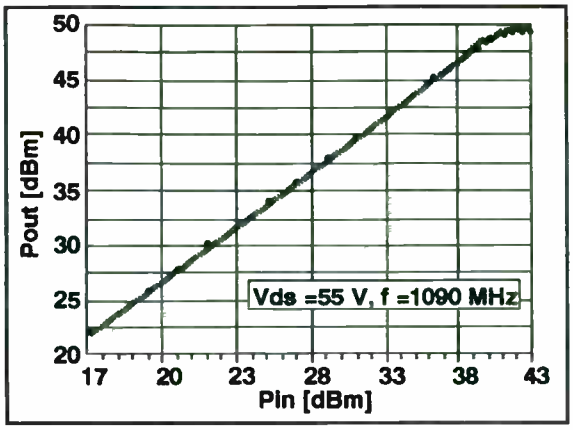


Figure 9. IFF CAM: Pout vs. Pin.

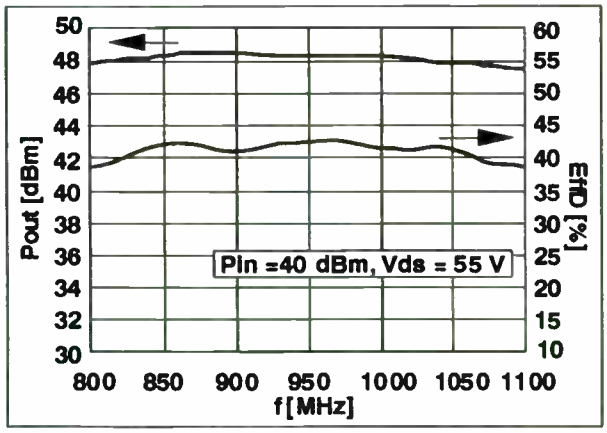


Figure 10. Rroadband 65 W CAM. Pout & Eff_D vs. frequency

As mentioned previously, one of the SST CAM unique characteristics is the gain control capability. This feature is illustrated in Figure 12, where the gain/attenuation data of a 75 W CAM is plotted as a function of the SST gate DC voltage. As the 1100 MHz input RF signal is kept at a constant power level of 39 dBm, the CAM power transfer factor changes from approximately 10 dB down to -20 dB. This 30 dB change is accomplished by a 8 V swing in Vgs. Additional gain control data is shown in Figure 13, where output power is plotted as a function of Vgs at various input power levels.

The SST dynamic range of gain control illustrated here significantly exceeds bipolar transistor performance for these types of applications. State of the art BJT JTIDS pulse power amplifiers have to rely on complicated, bulky, and expensive electronic switches to rapidly modulate the output power. The MOSFET, like the BJT, has a narrower gain control dynamic range than the SST. No

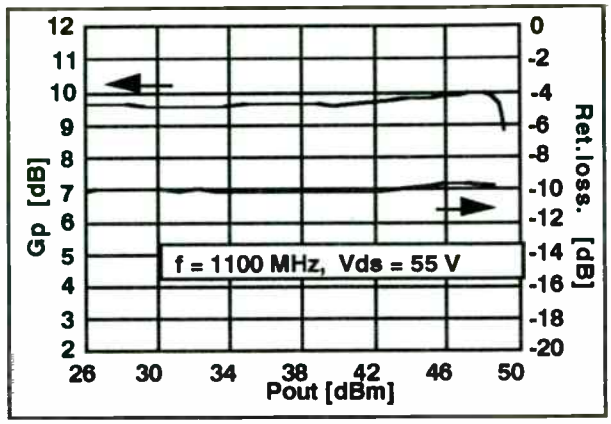


Figure 11. 50 W CAM: Gp & Input Return loss vs. Pout.

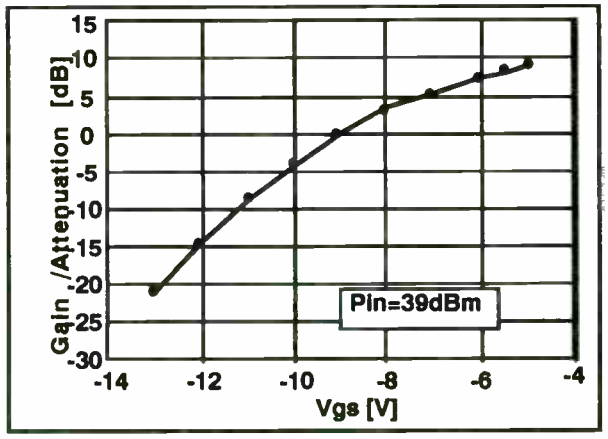


Figure 12. 75 W CAM: Power Gain/ Attenuation vs. Vgs.

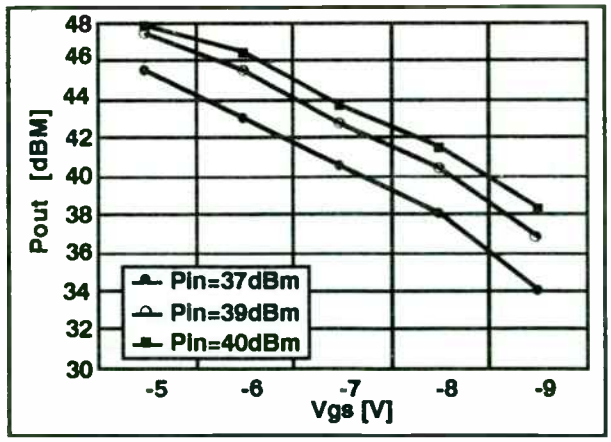


Figure 13. JTIDS CAM: Pout vs. Vgs f =1050 MHz, Vds = 55 V.

MOSFETs are known to be commercially available for comparison in this application.

In addition to varying the Vgs, one can also use the SST drain voltage to modify the CAM output power. Knowing the output power dependence upon the drain voltage helps one also define the power supply regulation requirements. Figure 14 shows the output power dependence upon Vds for a 200 W power CAM, tested at 800 MHz.

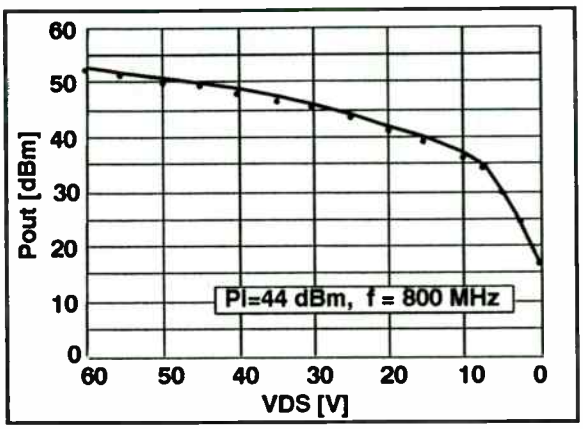


Figure 14. 100W CAM: Pout vs.Vds.

The 200 W CAM was fabricated by combining two 100 W CAMs using Wilkinson combiners. The basic CAM layout was shown in Figure 4. Figure 15 plots the output power and return loss data versus frequency for this CAM. Although the data was taken over the 700 MHz through 1200 MHz frequency range, this CAM was initially fabricated for 850 MHz - 1050 MHz operation.

Most of the CAM data illustrated so far is for modules set for class A operation. Data for a class B 100 W CAM tested at 800 MHz is shown in Figure 16. Notice the increase in efficiency as a consequence of the change in the SST operating point.

Another CAM characteristic mentioned above is the excellent pulse spectral response. A typical frequency domain pulse response for a 75 W CAM is shown in Figure 17. One can see the good symmetry of the sin(x)/x power spectrum lobes which confirms the low level of phase modulation generated by the SST during the pulse. To achieve similar results using BJTs, sophisticated pulse shaping circuitry operating at high currents would have been required.

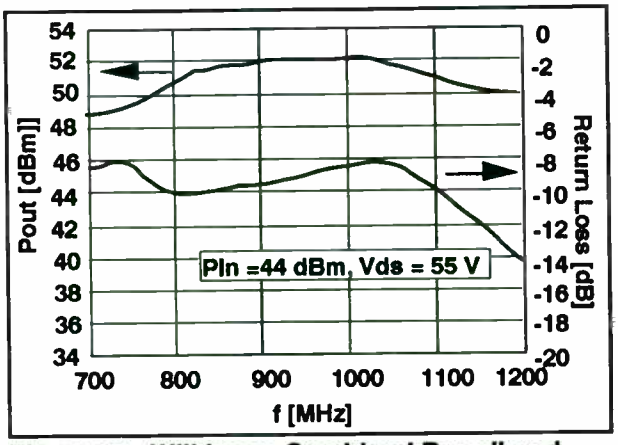


Figure 15. Wilkinson Combined Broadband 200 W CAM. Pout & Eff_D vs. frequency.

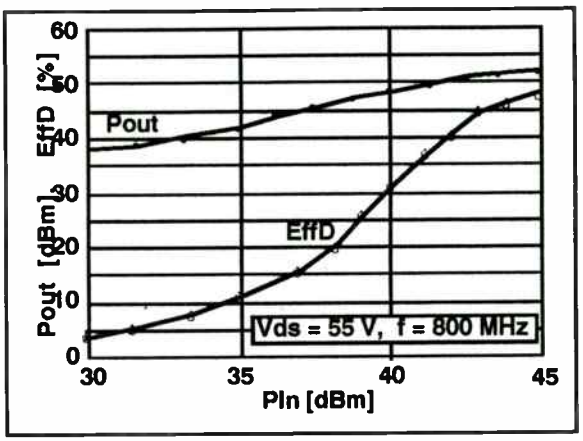


Figure 16. 100W CAM: Pout & Efficiency vs. Pin.

As a thermally stable device, the SST power capabilities show reduced sensitivity to temperature variations. Figure 18 contains peak output power data vs. the package flange temperature for an SST tested at 750 MHz and with a Vds of 55 V. The power derating factor resulting from this figure is 0.01 dB/°C.

The following list illustrates the power the MwT pulsed power SST CAM capabilities:

- Frequency: 960 MHz - 1215 MHz or 1030/1090 MHz (add 0.5dB)
- Pout: 50 W, 8.5 dB
  size: 1.75" x .75" x .25"

- Pout: 5 W, 8.5 dB
  size: 1.75" x .75" x .25"
- Pout: 100 W, 8 dB
  size: 1.75" x .75" x .25"
- Pout: 150 W, 7.5 dB
  Wilkinson combined
  size: 2.5" x 1.5" x .35"
- Pout: 200 W, 7 dB
  Wilkinson combined
  size: 2.5" x 1.5" x .35"

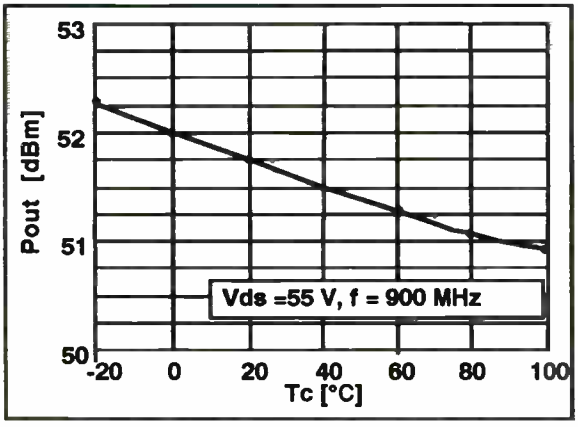


Figure 18. Power SST: Pout vs. Temperature.

Conclusions.

The development of hybrid circuits using the Solid State Triode is leading to a new capability for the designers of pulsed RF systems operating in the 800 to 1300 MHz region. Using Compact Amplifier Modules power amplifiers can be reduced in size and complexity. The internally matched CAMs can be easily combined to achieve almost any practical power level. The need for pulse conditioning and the use of special high power attenuators for power control can be eliminated saving precious real-estate in airborne and mobile systems.

Development of the SST CAM is continuing. Improved geometry SST chips and other biasing approaches will result in higher power and improved efficiency. New processing technology will increase the device cut-off frequency boasting the gain at lower frequencies and extending practical operation to near 2 GHz, further expanding the applications for the CAMs in radar, while CW operation at power levels of 10 W to 25 W can be applied to wireless digital communications systems.

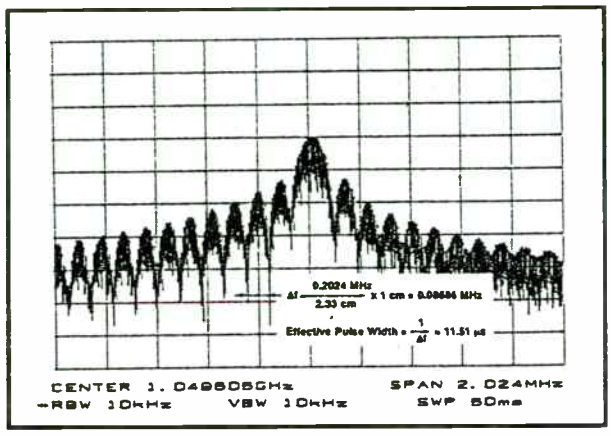


Figure17.RF Pulse Spectral Response.

Acknowledgements.

The SST and CAM work was supported in part by the Naval Air Warfare Center, Warminster, PA and by the Wright Laboratory, Dayton, OH.

The authors also acknowledge the critical contribution made at MwT by N. Thornton and P. Pacada, and the support and guidance provided by M. Omori.

Bibliography

A. Cogan, "SST for Microwave Applications", RF Expo East Proceedings, October 1989, Atlantic City, NJ.

S.McIntyre, "Demystifying New Generation Silicon Power FETs.", Microwave Journal, April, 1984.

H.O.Grandberg, "Building Push-Pull Multioctave VHF Power Amplifiers.", Microwaves & RF, Nov., 1987.

M. Abdolahian et. al. "Design and Performance of a Broad-band, High Voltage SIT Power Amplifier." RF Expo, East. Oct. 1991, Orlando, FL.

FREQUENCY SYNTHESIZER STRATEGIES FOR WIRELESS

Bar-Giora Goldberg, MSEE
Henry Eisenson
Sciteq Electronics, Inc.

PURPOSE

This presentation illustrated some of the challenges posed by emerging wireless/PCN/PCS industries (hereafter "wireless"), particularly as they apply to a key subsystem: the frequency synthesizer. Wireless creates a spectrum of opportunity and technical challenge, as well as risk, and since the frequency synthesizer is one of the most difficult developments in any analog system this document will seek to present some potential strategies for meeting those requirements.

Optimistically, the authors have written to two audiences. The first includes design and system engineers, who may be very competent at designing the RF systems of five years ago but with the advent of wireless are encountering new arrays of challenges based on economic, production engineering, and performance factors. The second target audience includes program managers and company executives, who may have limited technical knowledge yet must listen to the engineer's explanations as to why the strategies of yesterday simply don't work in this new technical and marketing environment.

DEFINITIONS

One of the problems facing many designers of systems and subsystems for the wireless businesses is a lack of structure. Operating bands, modulation schemes, protocols, power limitations, and – surprisingly – even applications, are all poorly defined. Firms developing systems and making investments in this new market, and particularly those who are aggressively working to establish a position early in the evolution of the industry, are very much at risk because of the lack of structure, protocols, spectrum allocations, and – of course – definitions. It is beyond the scope of this paper to attempt to generate such a set of data, however, so this statement is presented only as a warning to those investing money and engineering, and a plea to those involved in the establishment and standardization of protocols and definitions.

Regarding frequency synthesizers, for the purpose of this document historical definitions apply. Though they will not be listed in this paper, they correspond with those appearing in many publications.

INTRODUCTION

Radio systems are tuned by generating a frequency reference, and the quality of that reference determines the performance of the radio circuit. Accuracy is critical; both transmitters and receivers must be at the same frequency for communication to occur. When the energy of the signal appears only at the desired frequency, phase noise is considered perfect, and as that energy spreads to nearby frequency, phase noise is described as less perfect. A good reference generates no discrete uncommanded signals (spurs). Together, accuracy, phase noise and spurious signals define the performance of a reference.

By far, the best way to tune a radio or any RF system is with a crystal, and when multiple frequencies are necessary, multiple crystals and a switch can be used. Eventually, however, it becomes first impractical and then impossible to use complex arrays of crystals, as shown in Figure 1.

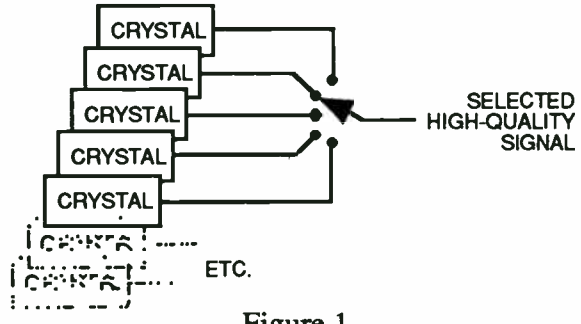


Figure 1

A "frequency synthesizer" is a device or circuit that synthesizes a new frequency based upon an original one (reference), retaining the stability, accuracy, and spectral purity of the reference though at a new point in the spectrum. It can generate one frequency (beyond that of the crystal reference) or multiple frequencies, as selected by a control mechanism (Figure 2).

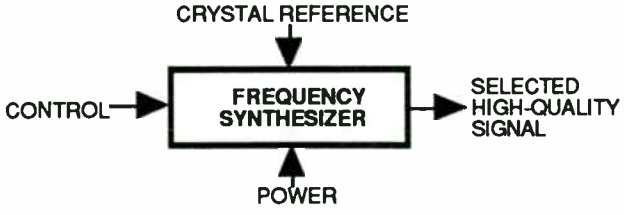


Figure 2

Like shoreline, spectrum is limited and precious, and the synthesizer (among other factors) determines how efficiently spectrum is used and how many channels can be compressed into any given operating band. Frequency synthesis is always a challenge, and the characteristics of the synthesizer have long defined the performance of the system that employs it.

Though there are as many synthesizer designs as there are designers, only three synthesizer techniques can be described as fundamental; all others are variations or combinations of one or more of them.

DIRECT-ANALOG

This (mix/filter/divide) is the oldest frequency synthesis method. The first time an engineer amplified a reference sufficiently to saturate a diode and used a filter to pick out a new frequency derived from the original, direct-analog was born. Today, many direct-analog techniques exist for multiplying, dividing, adding and subtracting an array of references, all locked to a common reference, to produce new frequencies.

This process supports very high spectral purity, since there is no correction circuit's "seeking" (that corrupts phase noise in a PLL), and careful planning can achieve frequency manipulation that avoids spurious signals. An important advantage of this technique is fast switching. The major drawback of direct-analog is the cost of the array of references required to cover the desired frequency range, plus the cost of one echelon of mix/filter/divide circuitry for each decade of resolution (step size) required.

Nevertheless, the finest synthesizer performance achieved by the industry employs direct-analog techniques, and appears in the Comstron FS-5000. Expensive, but there is no better method for generating a fast-switching, high spectral purity, broad bandwidth signal. Obviously, while this may be interesting to the designer of a simulator or radar system, the direct-analog approach is of little interest to a designer of wireless systems.

DDS

In one sense, the only true "synthesizer" is a DIRECT-DIGITAL SYNTHESIZER (DDS), since it literally constructs the waveform from the ground up (synthesizes the frequency) rather than combining or controlling existing oscillators.

A "vanilla" DDS appears in Figure 3, which also shows the signals generated by each circuit element.

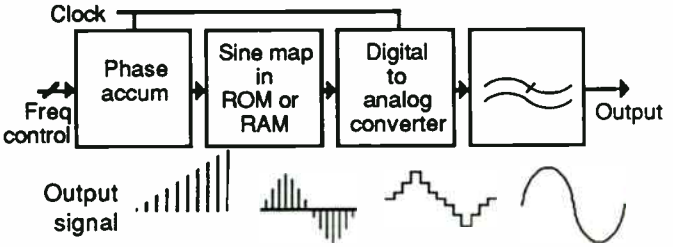


Figure 3

Though each of the blocks can be broken into many smaller ones (even down to the transistor level), the arrangement shown permits understanding of the means by which a DDS constructs a waveform.

The PHASE ACCUMULATOR correlates the clock with a control word, defining a "ramp" from 0° to 360° within some time period, and therefore the output frequency. A MEMORY maps the phase data in that ramp to a series of digital amplitude words, and a digital-to-analog converter DAC converts those digital data to an analog sinusoid (or any waveform stored in memory). That process must comply with sampled data rules, and therefore any frequency to be produced must be sampled *at least* twice each 360° (cycle), and the highest frequency that can be digitally generated is exactly one half the clock rate – though filter realities make the available output closer to 45%.

The DDS is also a supreme modulator. Digital shifting of frequency supports FM and toggling between two frequencies supports MSK/FSK; in fact, the DDS is a theoretically perfect FSK modulator. By placing an adder between the accumulator and the memory, the output can be shifted in time (phase), and a multiplier between the memory and the DAC permits scaling of the output, and therefore amplitude control. Both phase and amplitude modulation is therefore possible with a DDS, and with digital precision not possible with analog circuitry. Finally, SSB implemented using DDSs approaches theoretical perfection.

The advantages of the DDS include inexpensive high resolution (fine step size), fast switching speed, excellent phase noise, and while the signal is in the digital domain it can be manipulated/modulated with exceptional accuracy. The disadvantages include the fundamental limit of bandwidth (maximum frequency output is less than one half the clock rate, and logic has limits), and discrete spurious signals at a higher level than with other techniques. Nevertheless, in only twenty years the DDS has grown from an engineering novelty to a serious design tool. There are many DDS products on the market today, and they're generally divided into two categories

based upon a combination of price and performance. The "commodity DDS," typically in CMOS and from Analog Devices, Harris, Qualcomm, and Stanford Telecommunications, is characterized by low price, performance that is most often exploited as part of a much more complex synthesizer, and a high level of integration. Such DDS products often offer waveform manipulation capabilities, modulation, and other features sought by the wireless designer, but do not operate in the frequency ranges of most wireless systems.

High performance DDS products are often executed using very fast silicon or gallium arsenide logic so they can be clocked at a much higher rate than the commodity-level products. These DDS products cover a band much broader than that of the commodity DDSs. Expanded bandwidth requires higher clock rates, and therefore faster logic and more critical manufacturing and testing processes, hence higher prices. Even the fastest of the high performance DDSs operate only at about 400MHz, and that plus their relatively higher prices make the DDS unsuited to the needs of wireless systems.

While there are many specifications and architectures that differentiate one DDS from another, in general, a DDS can be characterized by a combination of bandwidth and spurious signals. By that simple standard, the state of today's DDS art will not meet the needs of wireless. The DDS should not be ignored by the wireless system designer, however, because as will be seen there are techniques by which the performance of the DDS can be translated to the frequency ranges of interest.

PLL

The PHASE LOCKED LOOP (PLL) is the single most commonly used synthesis technique, and is unquestionably the most flexible and adaptable. Perhaps more than ninety-nine percent of all synthesizers use one variant or another of the PLL. It appears in countless automobile radios and television sets, yet variants of the same architecture are used in exotic satellite transponders.

Figure 4

In Figure 4, a free-running oscillator generates a signal, the frequency of which varies (drifts) over time, according to circuit anomalies, temperature, etc.

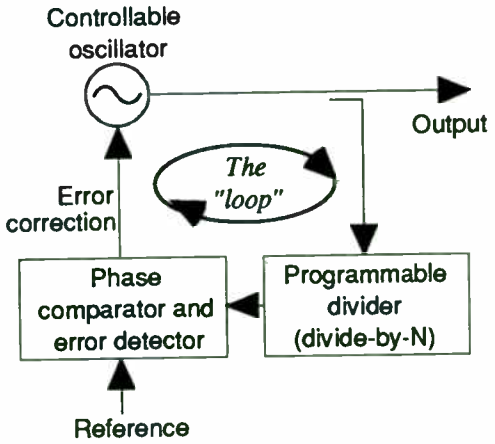


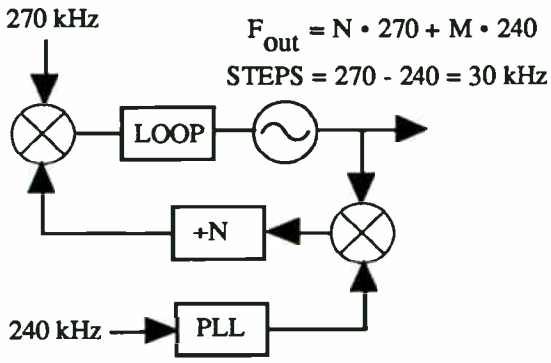
Figure 5

Figure 5 adds feedback and a phase detector, a correction mechanism that completes a loop to lock the output to some reference (thus "phase locked loop" or PLL), in accordance with some numeric ratio set by the frequency control command.

The PLL does not synthesize a waveform; rather, it controls the oscillator to the desired output frequency by dividing the output of the oscillator by some number, and then comparing the result with a reference. When errors are detected, they produce correction signals that return the oscillator to the correct operating frequency. For that reason, the system is always "seeking perfection," and the effectiveness of that seeking process determines the performance of the synthesizer. As that seeking occurs, the greater the deviation from the commanded frequency the worse the purity of the output signal. Such deviations are called "phase noise."

The advantages of PLL are low cost and excellent spurious suppression. The major disadvantage is that *both* fine steps *and* good phase noise can be achieved only in expensive implementations. In fact, the primary deficiency of PLL is that inverse relationship between step size and phase noise, because as step size decreases, division ratios in the system must increase, and the higher the division ratio the worse the phase noise within the loop bandwidth (close to the carrier). There is another drawback to PLL: switching speed. While the direct approach (DDS and direct analog) can be very fast, the PLL is slow because there is a certain electrical inertia involved before the system settles at a new frequency. The finer the required steps (the higher the division ratio) the longer it takes a typical PLL design to reach a new commanded frequency.

Phase noise can be reduced by using two PLLs wherein a primary loop generates the required operating band but in coarse frequency steps, therefore with low division ratios and acceptable phase noise. A second loop, also doing coarse steps, is combined with the first to generate fine steps (a difference). See Figure 6.



TYPICAL TWO-LOOP SYNTHESIZER

Figure 6

Switching speed can be improved by generating a tuning signal to the VCO early in the change process. That's usually done by generating a digital change word, and converting it to an analog voltage applied to pre-tune the VCO. This approach can increase speed, but it also increases circuitry, cost, power dissipation, etc.

Another approach is to simply use two PLLs, with a switch to select between them. Assuming only that the system "knows" the next frequency, it can tune PLL-2 to that frequency while PLL-1 dwells at the prior frequency. When the time comes to change, a digital command switches to PLL-2. This obviously requires two PLLs, with twice the power, etc.

As can be seen, the PLL is a very useful technique, but a conventional single PLL cannot achieve aggressive combinations of fine steps, fast switching, and good phase noise.

COMBINATION DESIGNS

Many systems combine two or more of the basic techniques, and this approach is constantly being explored for wireless applications. To cover a limited band in fine steps, above the range of a DDS, the output of a DDS can be mixed with an LO and a filter used to select the desired sideband. In direct-analog systems, a PLL might be used to generate some of the required references. PLLs are often successfully combined with DDS to achieve fine frequency steps with reasonable phase noise. Ever more inventive combination designs appear every year, and some include elements of all three building blocks: PLL, DDS, and direct analog.

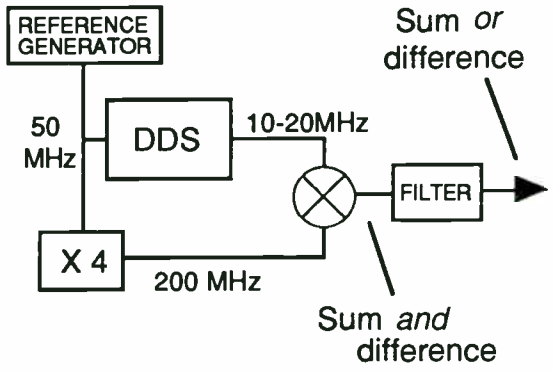
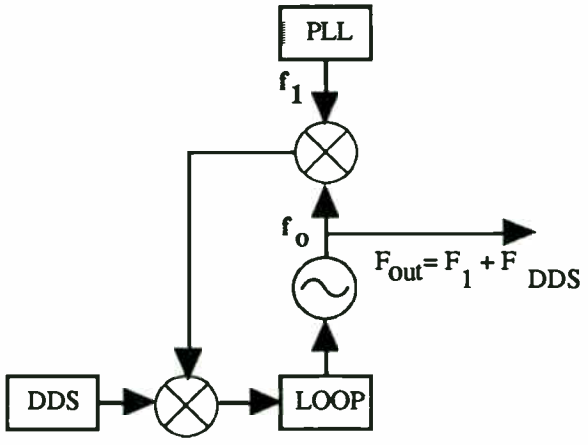


Figure 7

When a system concept dictates operation over a relatively narrow band but at a higher center frequency, a DDS can be upconverted to that range so as to exploit the DDS' fine steps, modulation capabilities, and fast switching. An example of such an architecture appears in Figure 7. In the example shown, 10 MHz of the DDS' output is translated or upconverted to the 200 MHz range. Obviously, the same architecture can be used to achieve a wide variety of goals, limited by DDS bandwidth and the physics of the filter.

It is sometimes useful to upconvert the DDS with a PLL, and, again, there are many ways to accomplish this. One mechanism (developed in 1984) uses a DDS as a reference to the phase detector of the PLL, though multiplication seriously degrades spurious performance.



TYPICAL DDS+PLL SYNTHESIZER

Figure 8

Another (also from that period) uses a DDS, a PLL, plus a combining loop. A typical DDS+PLL

approach is shown in Figure 8. These techniques are useful, but involve cost and power compromises.

There are other interesting PLL mechanisms, which will be discussed as the needs of the emerging wireless market are analyzed.

GENERAL REQUIREMENTS OF WIRELESS SYSTEMS

The wireless industry will be filled with cell or base stations, of which each interacts wirelessly with dozens, hundreds, or even thousands of portable stations. The consumer purchases the wireless portable but the price of the cell/base radio is usually buried in service charges by the network owner. Also, because power, cost, and size are less important in immobile cell/base radios, this paper addresses only the requirements of portable systems.

It also considers only frequency-agile systems. The synthesizer for a typical wireless system (for instance a TDMA-based wireless PBX) will operate near 900 MHz with steps under 50 kHz, spectral purity that supports digital modulation, and power dissipation that ensures reasonable battery life. CDMA-based designs can trade frequency agility for complex coding circuitry, but TDMA, N-AMPS, AMPS, GSM, and most other cellular/PCS/PCN wireless systems require channel selection, hence agility.

Throughout wireless, once the basic function is accomplished price becomes paramount. Few engineers are experienced in designing products for the volumes involved in the wireless market. When millions of units will be made, savings of a few cents on each can have a profound effect upon the profitability of the enterprise. There is a spectrum of companies addressing every emerging opportunity in wireless, hence competition is fierce and will only grow tougher as the industry evolves. The relationship between price and performance, therefore, determines corporate success and failure.

Required performance is established by marketers who determine what the end user needs and wants, and what limits are acceptable. Once that information is collected and integrated, it is the engineer's task to achieve that performance at the lowest possible price. The most general assumption is that parts count will affect price, particularly when assembly effort is considered, and therefore the simplest (or more highly integrated) product will probably be lowest in cost and most competitive.

In wireless, reliability and durability are only barely behind price in relative importance. Historically, our industry has defined quality as either "military" or "commercial," and the assumption was that "military quality" produced the highest reliability. There are two reasons why production methodologies that define "military" quality won't work for wireless. Military quality standards require training programs, a wall full of diplomas, constant inspections, intensive testing, and reams of documentation. Like the ubiquitous pager, wireless products are measured only by the simplest standard of all; they must work nearly forever.

Generally, in the electronics industry, parts count and reliability are inversely related, and the more parts the more opportunities for manufacturing defects, so the simplest (or more highly integrated) product will be the most reliable.

Size is an issue because wireless system developers are driven toward exceptional portability, and small size always complicates RF/analog design problems, involves higher levels of integration, and generally raises the cost of engineering the final product.

Some system architects have described requirements that include rapid settling. GSM (the European standard), for instance, establishes timing standards that dictate a fast switching solution. Designers of PBX and security systems have also asked for fast switching. A generic wireless synthesizer solution must, therefore, include the ability to switch at rapid rates. By extension, that seems to demand either multiple synthesizers with a selector and supporting software, or some implementation of a direct design.

"Wireless" implies exactly that, so not only is the communication by radio rather than coaxial cable, there are no extension cords and portable elements of the system run on batteries. Power consumption determines battery life, and one of the major engineering challenges of the wireless industry is efficiency of power utilization.

Marketing defines the needs of the end user and ensures that the product will have salable competitive advantages. The engineering group must develop a product that first attains those goals, but also costs little to make and works forever.

The needs of the wireless industry, in general, can best be met by simple, highly integrated, physically small, low-power designs. That is true in general, and it is especially valid for the frequency synthesizer engine that drives the system.

FREQUENCY SYNTHESIS FOR WIRELESS

The frequency synthesizer subsystem of a wireless product does exactly the same job as a synthesizer does for any RF system. It is the system's tuning mechanism, and its performance determines the product's compliance with many marketing-driven requirements. Channel spacing determines step size, and in a PLL that establishes a mathematical relationship between operating band and division ratios that also dictates phase noise performance. Spurious signals can fool the system, lack of reliability can kill it, and excess cost makes everything else insignificant. This is valid for all RF systems and subsystems, including the synthesizer.

We've learned that the frequency synthesizer "menu" offers many alternatives from which the system designer can select, and we've also covered some of the needs of marketable wireless systems. Perhaps one valid method for identifying the best approach for wireless is to eliminate those techniques that will not work. The problem would be solved if a wireless portable end-user would pay for – and carry around – a Comstron FS-5000, or if the industry knew how to clock a cheap DDS at 2 GHz and still get good spurious suppression from the result. For reasons of cost, bandwidth, size, or spectral purity, therefore, pure-direct (analog or digital) synthesizers do not work for wireless.

What about a DDS upconverted using mix/filter circuitry – in other words, one of the "combination" designs? One advantage is the ability of the DDS to generate a modulated signal derived from digital manipulation of the waveform, and a second advantage is switching speed. At first glance, this approach looks inviting, but what are the disadvantages?

Make the estimate. Assume that the system demands 10 MHz bandwidth at a final operating frequency near 900 MHz. If reasonable filters are to be used, then there must be either two upconversions or substantial bandwidth in baseband. Two upconversions are expensive, since two good LOs must be generated and that greatly increases complexity.

For a single upconversion the required baseband coverage might be 50 MHz, which implies a clock rate for the logic near 110 MHz. Consider Figure 9, which illustrates the limitations imposed by filter issues.

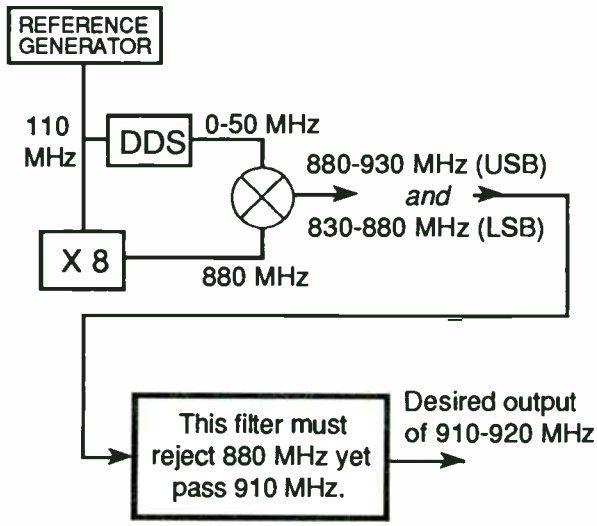


Figure 9

Depending upon the required suppression of the undesired sideband, the filter requirements shown are not impossible to meet. They are, however, difficult, very expensive, labor intensive, and impractical for inexpensive portable equipment.

Power is another issue that probably disqualifies the DDS. Even an advanced-technology, fractional-micron CMOS, fully integrated DDS will occupy a die of about 100 x 100 mils, and assuming a clock rate of 100 MHz the most optimistic dissipation estimate is 0.5W. Wireless portables will probably use a tenth of that for the entire design, hence the likely disqualification of DDSs using today's technology.

For these reasons, neither two conversions nor a 100 MHz DDS clock rate are practical. The designer must look beyond the direct approach. That leaves indirect, or PLL, which works for nearly all RF applications throughout the electronics industry.

In today's wireless, synthesizers are almost all PLL and use either one or two loops. Both have been implemented with very high levels of integration, and single chip PLL ASICs are available at $5 or less, with operation from 6-15 mA. These devices include all functions and require only an external VCO plus a few discrete devices for the loop filter.

The single loop iteration of such a design is simple, economical, reliable, and small, but it does not meet the requirements of digital modulation techniques.

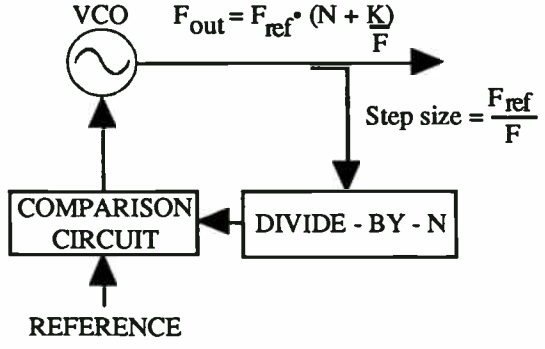
For many years, a broad variety of simple single loop synthesizers have used inexpensive PLL chips by Fujitsu, Plessey, Motorola, etc. to meet the requirement of early wireless, including FM/analog

cellular. Close channel spacing of agile wireless requires fine tuning steps and therefore high division ratios, yet digital modulation demands phase noise better than conventional PLL architectures can achieve.

Designers have therefore used two loops to achieve the combination of phase noise and step size needed by digital modulation schemes, which means two synthesizers with two VCOs, etc., and all the baggage a double circuit implies. The two-loop solution therefore implies higher power, complexity, cost, and likelihood of failure – everything the successful wireless system designer was directed to avoid. Nevertheless, state of the PLL art for wireless has been two loops,which achieve the required phase noise. In some systems, the requirement is for two such assemblies with a switch, to support fast switching. That's four loops, with four VCOs and associated circuitry; it's expensive, large, power hungry, and in general is an array of compromises.

FRACTIONAL *N*

There's a relatively little-known derivative of PLL that can address many such requirements with only one loop; it's called fractional *n*. Though mysterious to many engineers, it is not a true innovation because some fractional designs have been used for years. Fractional *n* synthesis can best be understood by examining the conventional block diagram for a PLL, but with one twist. The divide-by-N circuit appears where it always has, but additional circuitry changes the *value* of N from an integer to some fraction. Consider the impact, as suggested by Figure 10.



Conventional PLL: N is an integer
Fractional: N can be a fraction

Figure 10

N is programmable, of course, to define the relationship between the output and the reference. Suppose the requirement is to operate at about 1 GHz in 30 kHz steps. In conventional PLL designs, the output (1000 MHz) is divided by 33,000 to reach a value that can be compared to the reference (30 kHz), and that division imposes a phase noise degradation of 20 log *N*, or 90 dB. That degradation can be cut by 20 dB with a good fractional design, and there are several ways to achieve the desired result. Fractional division can be used to *increase* the reference by an order of magnitude, or to *reduce* the division ratios, and in any case the effect is to reduce the division ratio for a given combination of output frequency and step size, thus improving phase noise.

Fractional at first appears to be an aspirin for all forms of PLL headaches, but that hasn't been the case. Conventional implementations of fractional require major increases in circuitry, and that implies increased power dissipation, parts count, complexity, labor costs, and therefore overall costs – again, all the things the designer was told to avoid.

Naturally, fractional division generates a new periodicity within the divide-by-N circuitry, and therefore introduces spurious signals. For the above example, if the reference goes from 30 to 300 kHz, for 30 kHz steps there will be 30kHz/60/90...etc. spurs in addition to the 300 kHz spurs. Those spurs are at the frequency of the new periodicity or its harmonics, but in either case they can produce a major degradation of spurious suppression. In a way, fractional has been largely ignored because those who have experimented with it consider the improvement in phase noise to be virtually a trade off for spurs.

Yet another problem with fractional is that the fractionality of conventional circuits is fixed, hence a given design has little flexibility and new applications imply new designs. For these and allied reasons, fractional *n* synthesis has been largely ignored. This situation is about to change, as market factors drive engineering to examine every alternative.

An important derivative of fractional is Sciteq's Arithmetically Locked Loop (ALL), which combines aggressive digital signal processing (DSP) with fractional to overcome all of the disadvantages outlined above. ALL is a simple – and patented – design that increases overall gate count (compared, for instance, to the Fujitsu or Plessey PLL chips) by about 3%, so complexity is not an obstacle. The effect of the new periodicity is minimized, hence spurious signal level is also not an obstacle. Finally, fractionality is programmable, making one design suitable to a spectrum of applications.

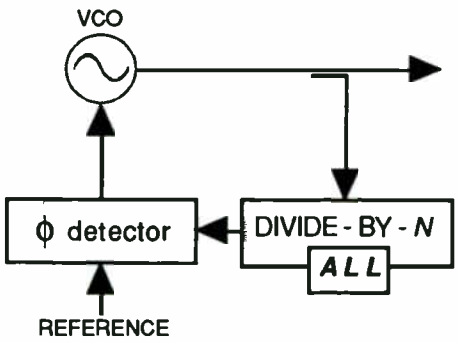


Figure 11

Adding ALL to a synthesizer design is simple, as implied by Figure 11. There is no change to the existing design except for I/Os in the divider circuit; the ALL block adds simple DSP that does the actual job of managing the division process and reducing the required ratios. In most cases, ALL will add approximately 3% more gates to non-ALL circuitry.

ALL APPLICATIONS IN WIRELESS

The ALL can operate over all wireless segments of the spectrum up to C-band. Comparisons are best made by examining any of the standard single-chip solutions (Fujitsu, National Semiconductor, Plessey, etc.) and adding about 3% in gate count and power while reducing close-in phase noise by about 20 dB.

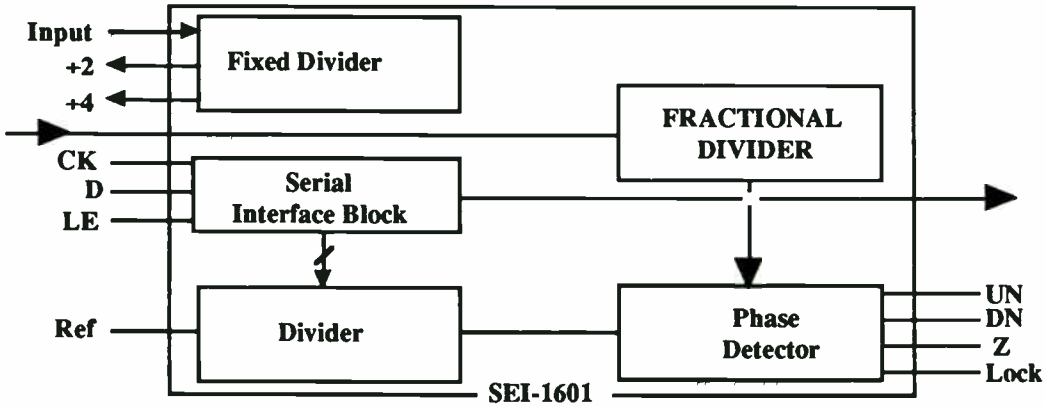
That performance level support digital modulation scenarios with a single loop solution. Even close-in, phase noise is competitive with levels ordinarily achieved using two loop designs.

Frequency switching is faster than conventional PLL designs, even without pretuning, since the reference frequency can be much higher for a given step size.

ALL has been in full production for two years, executed with discrete components to build modular synthesizer products. Potential ALL users can today support system development using prototypes built with discrete devices, in the expectation that a fully integrated solution will be available late 1993.

Many technologists who have evaluated ALL believe that some derivative will appear in most successful cellular and other wireless products that require frequency agility. Developers of such systems should contact Sciteq.

Arithmetically Locked Loop as a 2.5 GHz monolithic synthesizer:



- **20 dB phase noise improvement over other single loop designs**
- **Integrated +16/17 dual modulus**
- **2.5 GHz front end**

Time Division Multiple Access (TDMA) Transmitters: Characterizing Power, Timing, and Modulation Accuracy

Helen Chen
Hewlett Packard Corporation
Microwave Instrument Division
1400 Fountain Grove Parkway
Santa Rosa, CA 95403

This paper explores the test requirements of digital RF communication transmitters, presenting practical measurement techniques that reduce system development time and installation costs. Practical spectrum analyzer measurements for TDMA carriers are described, including power, timing, and modulation accuracy tests. Examples from the NADC, PDC, GSM, and CT-2 systems are presented.

The new digital TDMA communication systems present diverse and complex test and measurement issues. Each digital RF system has a unique set of measures for transmission that ensure high quality signals. This paper relates some of the quality metrics to characteristics of the TDMA signal under test. We examine the spectrum analyzer as a powerful, flexible solution to TDMA transmitter measurement challenges. Many digital TDMA transmitter tests can be done with new, expanded spectrum analyzer measurement capabilities. We describe practical spectrum analyzer measurement techniques for TDMA carriers, including carrier power, carrier-off power, adjacent channel power, spurious emissions, and TDMA burst timing. A new digital RF demodulation capability is examined for extracting digital modulation metrics such as error vector magnitude and I/Q offset. Measurement optimization and performance limits are described, and examples from the North American Digital Cellular (NADC) system are shown. We also consider other digital RF formats, including GSM, DCS-1800, PDC, CT-2, and DECT.

Key TDMA Measurement Challenges

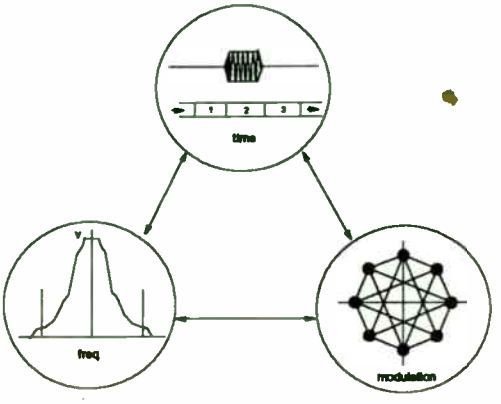


Figure 1: Digital TDMA Signals

Digital RF communication systems have complex, many layered signals. Performance requirements must be met in three domains: the time domain, the frequency domain, and the modulation domain. A time division multiple access (TDMA) channel structure increases carrying capacity by placing multiple users on the same carrier frequency, each user transmitting in turn on the carrier. The user's turn, or time slot, must be precisely placed in time to avoid overlaying users in adjacent time slots. The transmission must comply with a strict time and power mask. In the North American Digital Cellular (NADC) system a mobile time slot is 6.67 ms long.

In the frequency domain, each channel is restricted to a narrow bandwidth, and therefore requires high spectral efficiency. Spectral interference in adjacent channels has many causes, such as poor power control or problems with digital modulation. The NADC channel bandwidth is 30 kHz. Adjacent channel power and spurious emissions are carefully specified in system standard documents.

The quality of digital modulation must also be assured. Clear, error-free transmission is essential to voice and data

quality. Digital modulation metrics are tailored to each modulation format. In the European Global System for Mobile Communication (GSM), the global phase of the Gaussian minimum phase shift keyed (GMSK) modulated carrier must be less than 5 degrees RMS. In the NADC system, the pi/4 shifted differentially encoded quadrature phase shift keyed (pi/4 DQPSK) modulated carrier must have an error vector magnitude less than 12.5 percent RMS.

The complex signals of a digital TDMA system are required to support high capacity, high performance service at a low cost. In addition, digital TDMA is a new technology with rapidly changing system structure and modulation formats. Digital TDMA systems are proliferating worldwide. Pacific Digital Cellular (PDC), European Global System for Mobile Communication (GSM), and North American Digital Cellular (NADC) are spreading rapidly, an evolving to fit their markets. Assuring the quality and the low cost operability of these systems requires flexible, low cost test equipment tailored to the new digital TDMA performance metrics.

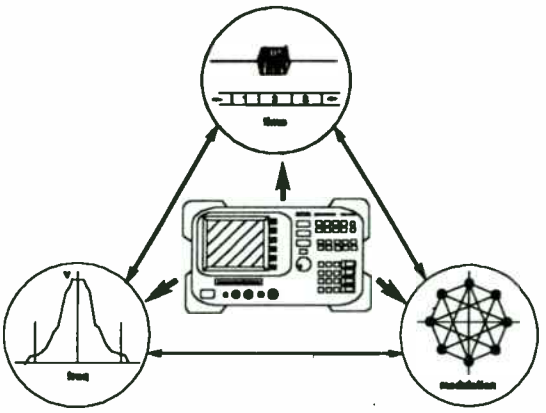


Figure 2: Low Cost, Flexible Tools

A spectrum analyzer can be a low cost platform for test solutions for digital TDMA systems. The Hewlett-Packard 8590 E-series spectrum analyzer can be adapted to make many digital communication system measurements. The architecture supports upgrades to new formats in the future. An expanding set of custom options adapt the analyzer for time, frequency, and modulation domain measurement. On a spectrum analyzer platform, digital RF tests can be done over the full frequency range (9 kHz-26 GHz) and signal sensitivity of the instrument.

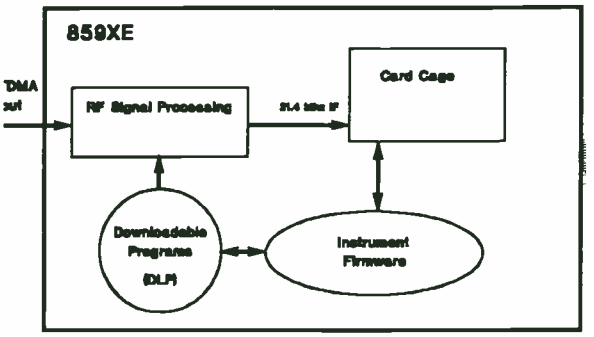


Figure 3: 8590E-Series Flexible Architecture

The HP 8590 E-series spectrum analyzers have a unique, flexible architecture that allows new measurement capabilities to be added quickly to the instrument. See Figure 3 . A traditional RF signal processing chain downconverts the carrier, and filters it through resolution bandwidth filters for subsequent power detection. The HP 8590 E-series instruments add an option card cage to the RF chain. This allows optional hardware to be connected to the signal processing, including extensions for precise time domain and modulation measurements.

The spectrum analyzer not only has flexible hardware, but also has adaptable software, in the form of downloadable programs (DLPs) on memory cards. (Downloadable programs are also called measurement personalities.) New measurement algorithms can be driven from code loaded from memory cards into the spectrum analyzer's memory. Spectrum analyzer hardware setup, trace manipulation, and test limits can be automated with a DLP. By changing the DLP loaded in the instrument, a GSM analyzer can be reconfigured easily into an NADC analyzer.

The key parts in adapting the spectrum analyzer will be highlighted as we describe practical measurement techniques.

Catching Time Conflicts

We will first describe measurement techniques using a spectrum analyzer to catch time conflicts in a TDMA system.

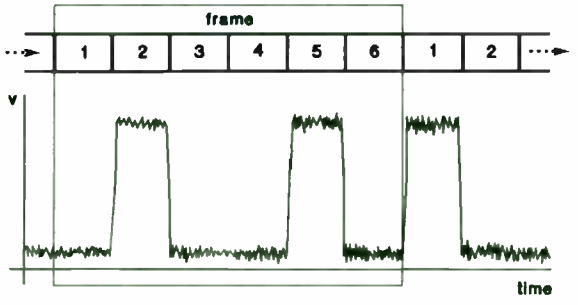
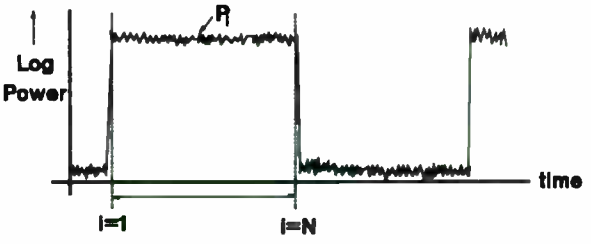


Figure 4: Burst Amplitude Modulation for TDMA

A time division multiple access (TDMA) system requires transmissions to be pulse-amplitude-modulated, or burst. Several users are time-multiplexed onto the same carrier to increase spectral efficiency. As shown in Figure 4, time on a carrier frequency is divided into a stream of frames. Each frame contains a fixed number of time slots. The time slot sequence repeats as the frames are transmitted. In the NADC system, a 40 ms frame had six 6.67 ms time slots. A user may transmit a packet, or burst, of modulated carrier only when his time slot appears in a frame.

In a TDMA burst carrier, good transmission power versus time characteristics are critical to avoid interference with adjacent channels in the time domain and in the frequency domain. Four key area of concern are carrier power, carrier off power, time position of the burst, and burst shape. Measuring these parameters is a good way of exposing potential interference problems.



$$P_{mean} = 10 \times \log\left[\frac{1}{N} \times \sum_{i=1}^{N} 10^{\frac{P_i}{10}}\right]$$

Figure 5: Carrier Power

where:

Pmean = mean carrier power during on part of burst (dBm)
Pi = power level at sample point i of the waveform (dBm)
n = number of sample points in the "on" part of the burst

Carrier power is defined to be the mean power transmitted during the active, or "on," portion of the burst. Testing the carrier power in the "on" portion of the burst ensures that burst power is sufficient for clear reception.

With a spectrum analyzer, the measurement is made with the instrument fixed tuned to a single frequency, yielding a power versus time trace on screen. The resolution bandwidth and video bandwidth in the RF signal processing chain are set wider than the channel bandwidth to avoid distorting the pulse shape. The trace data is averaged for the "on" portion of the burst.

Since the trace is defined on a logarithmic power scale, a true power average can be obtained only by performing an anti-log transformation to linear units prior to averaging the data. In general, the mean of log powers is not equal to the log of the mean power. In the equation above, each sample of trace data in the "on" portion of the burst (Pi) is anti-logged and summed. An average value is calculated using this linear unit sum. The mean power is then returned to the logarithmic scale for display as Pmean. For modulation formats with significant amplitude modulation, such as pi/4 DQPSK, true power averaging is essential to accurately finding the mean carrier power.

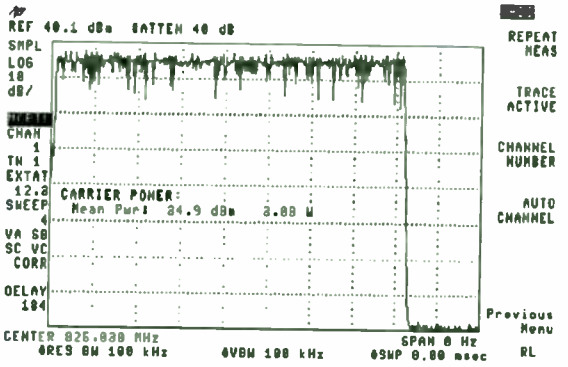


Figure 6: NADC Carrier Power

Here is an example of carrier power measured for a burst NADC-TDMA carrier. The HP 85718A NADC-TDMA downloadable program automates spectrum analyzer setup and power calculations.

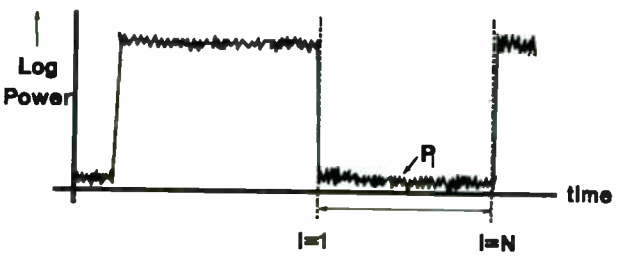


Figure 7: Carrier Off Power

Carrier off power is defined similarly to carrier power. Carrier off power is the mean transmitter power during the part of the frame where the transmitter is inactive or "off." In Figure 7, the inactive portion of the burst is in samples i=1 to N. The power in "off" portion of the burst must be low to avoid obscuring a burst in the other time slots.

The spectrum analyzer is again used in a mode fixed tuned to the carrier frequency. For better noise rejection and improved sensitivity, a narrower resolution bandwidth and video bandwidth are used to measure the low level "off" signal. The trace data is averaged for the "off" portion of the signal shown on screen.

The narrow video bandwidth averages the "off" power of the carrier, so that a stable power level can be read.

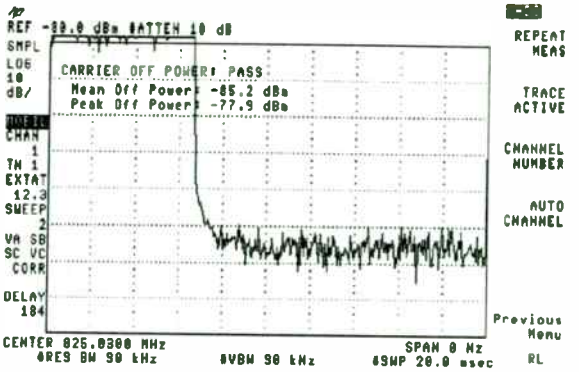


Figure 8: NADC Carrier Off Power

Here is an example of an NADC-TDMA carrier off power measurement, part of the HP 85718A NADC downloadable personality.

Note the 30 kHz resolution bandwidth used in carrier off power compared to the 100 kHz resolution bandwidth used in the carrier power measurement example. The results are compared to limits specified in the NADC standard document, IS-55. A "PASS" or "FAIL" indicator is displayed.

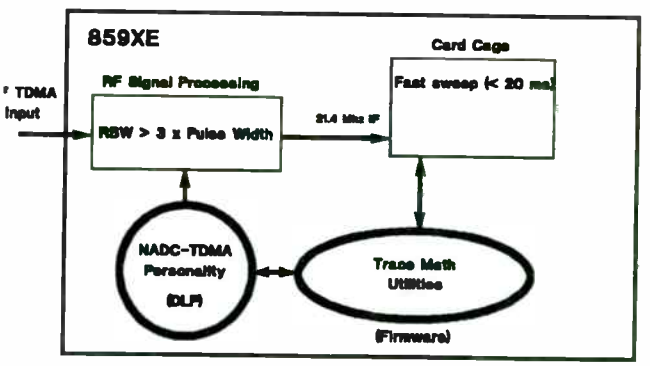


Figure 9: Carrier Power Measurement Tool

The HP 8590 E-series spectrum analyzer supports the tailored trace manipulation needed for true power averaging. The HP 85718A NADC-TDMA downloadable program automates a measurement customized for NADC test limits.

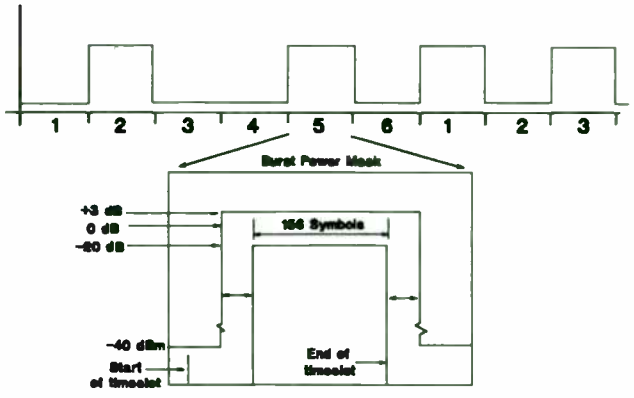


Figure 10: TDMA Burst Position in Time

Many standards define time domain masks for the position and shape of a TDMA burst., such as the one in Figure 10. Bursts should be verified to be in the assigned time slot. Transmission power must be maintained in a narrow window during the "on" part of the burst. The ramp-up and ramp-down times do not extend into adjacent time slots. Examining burst ramp shape can expose not only problems due to infringement on adjacent time slots, but also interference due to excess spectral splatter of burst power transitions.

To make the time position and shape measurements, the spectrum analyzer must be able to find and focus on any desired time slot. The transmission is then shown on screen and compared with the power and time limits specified in the standards.

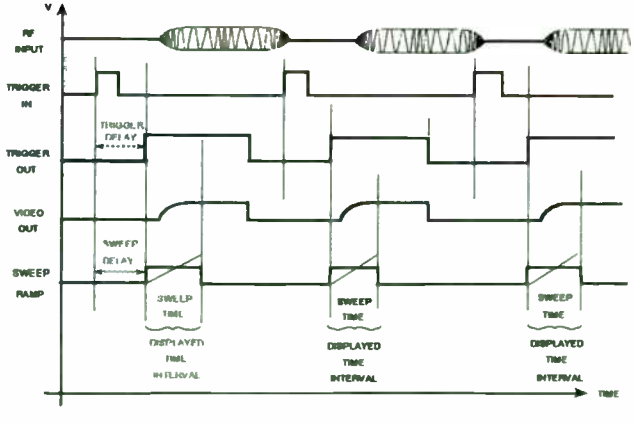


Figure 11: Delayed Sweep Trigger Timing

The HP 8590 E-series spectrum analyzers use an adjustable delayed trigger to focus the spectrum analyzer measurement sweep on the desired pulse. The coordination of this measurement is shown in Figure 11. At the top of the timing diagram, an RF burst is the input to the spectrum analyzer. Additionally, a TTL trigger input is required, usually synchronous with the frame rate of the TDMA RF signal. This input trigger is processed through hardware on the Option 105 time gate card , which provides a user adjustable delay of the input trigger signal. Using the time gate, the trigger output is positioned to be just before the time slot of interest. The delayed trigger is connected to the external trigger of the spectrum analyzer, initiating a sweep at the rising edge of the delayed trigger at the time gate trigger output. The spectrum analyzer sweep time is adjusted to capture the time interval of interest. In Figure 11, the timing diagram shows the capture of the rising edge of the burst.

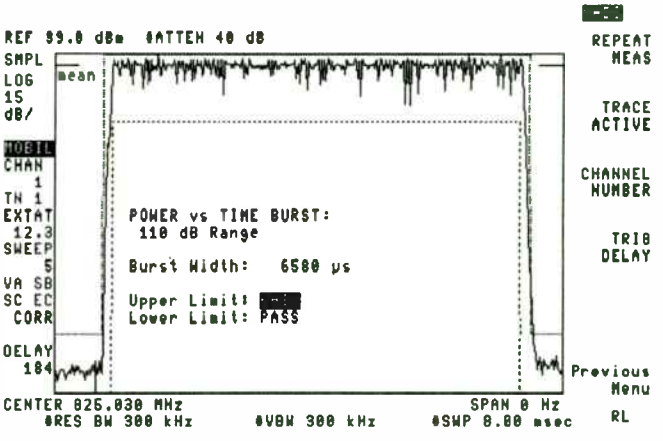


Figure 12: NADC Full Timeslot

The figure above shows a power versus time measurement executed on an NADC burst in time slot 1. This burst is barely within specification. Spectrum analyzer setup, measurement, and result display are completely automated in the HP 85718A NADC-TDMA personality.

Although spectrum analyzers are high dynamic range instruments, the signal range requirements for some TDMA system are a challenge. For example, the NADC-TDMA IS-55 standard specifies a -60 dBm burst "off" level. For this limit, a 3-watt mobile phone requires 95 dB of measurement range. The downloadable software can extend the on-screen calibrated dynamic range. Notice in Figure 12 that the screen displays 110 dB of range on

an analyzer that normally has an 80 dB log display. The personality controls the spectrum to analyzer take two sweeps: one optimized to measure the active portion of the burst, the second optimized to measure the inactive portion of the burst. The two sweeps are time matched, and pasted together. They can be displayed as one trace, showing 110 dB of display range.

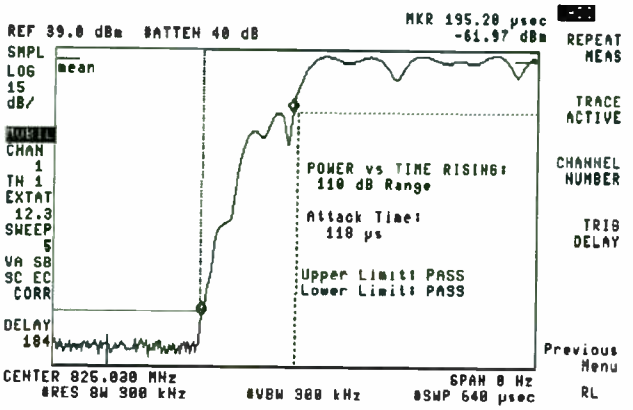


Figure 13: NADC Burst Rising Edge

In this example, the spectrum analyzer delayed trigger and sweep time have been adjusted to focus on the rising edge of an NADC burst. The time selection adjustments are automated in a DLP. A fast digitizer in the card cage is needed to extend spectrum analyzer sweep times to sweeps faster than 20 ms.

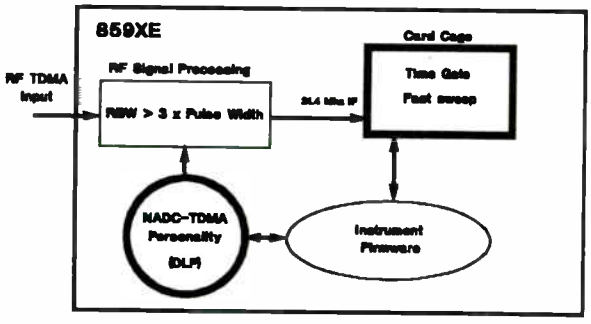


Figure 14: Burst Position Measurement Tool

The key additions to the HP 8590 series spectrum analyzer for digital TDMA time domain measurements include card cage hardware and downloadable software. The time gate and fast digitizer are needed to find and focus on the time intervals of interest. Adjustments are simple, if driven from the specialized DLPs.

Detecting Sources of Frequency Interference

In addition to time domain conflicts, frequency domain interference must be detected in a digital communication system. Both interference close to a carrier transmission and distant spurious should be kept at low levels.

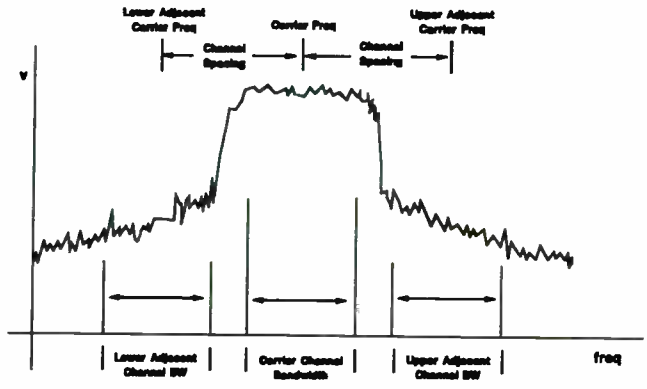
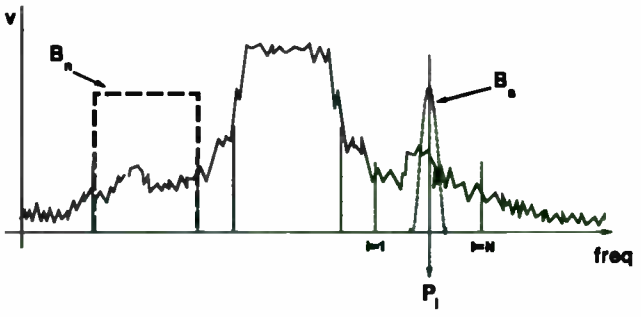


Figure 15: Adjacent Channel Power

Adjacent channel power (ACP) is a measure of the amount of leakage power spilled into adjacent channels. Usually, ACP is specified as a ratio to the total carrier power, but some standards specify absolute power levels. The total power in the lower adjacent channel or in the upper adjacent channel is compared to the total carrier power. In a mobile communication network, the interference from the adjacent channel must be kept very low. Users on one channel should be able to communicate without interference from another channel. High adjacent channel power can degrade communication quality or increase the interference rejection required in the mobile station for good overall performance.

In a spectrum analyzer, adjacent channel power measurement begins with obtaining a frequency domain trace of the spectrum, centered about the carrier channel of interest. The analyzer sample detector is used to accurately detect the power of the digitally modulated signal. The instrument resolution bandwidth is set to be much narrower than the channel bandwidth to yield a more accurate integration of power over the adjacent channel bandwidth. The video bandwidth should be at least ten times larger than the resolution bandwidth to remove any video averaging of the trace sample points. Video averaging can lead to errors as great as 2.51 dB in the sampled power reading.



$$P_{ACP} = \frac{B_s}{B_n} \times \frac{1}{N} \times \sum_{i=1}^{N} P_i$$

Figure 16: Continuous Carrier ACP

where:

Pacp = RMS-like power in the specified integration bandwidth (watts)
Pi = power level at sample i of the spectrum analyzer trace (watts)
Bs = specified integration bandwidth for adjacent channel (Hz)
Bn = effective noise bandwidth of spectrum analyzer (Hz) 1.13 x resolution bandwidth of spectrum analyzer
N = number of sample points in specified bandwidth

For an analog communications system or a continuous (non-burst) digital carrier, the ACP calculation is made after obtaining a frequency domain trace of the spectrum. Adjacent channel power is calculated in the spectrum analyzer using an integration method. The leakage power in the adjacent channel is computed with the power integration equation shown in Figure 16, above. The trace sample points (Pi) within the adjacent channel integration bandwidth are averaged in linear units (watts). The adjacent channel integration bandwidth is sometimes called the specified bandwidth. A scaling factor, Bs/Bn, is applied to compensate for the shape of the spectrum analyzer signal processing filters, the resolution bandwidths.

The power at each trace sample point is detected after filtering by the spectrum analyzer resolution bandwidth. The HP 8590 series resolution bandwidth is a four-pole synchronously tuned filter. This filter captures more power under its flared skirts than a straight-sided, rectangular filter of the same 3 dB bandwidth. The power measured by the spectrum analyzer must be corrected for the excess power caught under the skirt of the synchronously tuned filter. The factor Bs/Bn multiplies the average power in the adjacent channel by the number of effective noise bandwidths in the specified bandwidth. A rectangular filter correction factor is usually required, but NADC-TDMA standard IS-55 and IS-56 specify the application of a Nyquist square-root raised cosine filter prior to performing the integration. The HP 85718A NADC personality software applies the Nyquist correction.

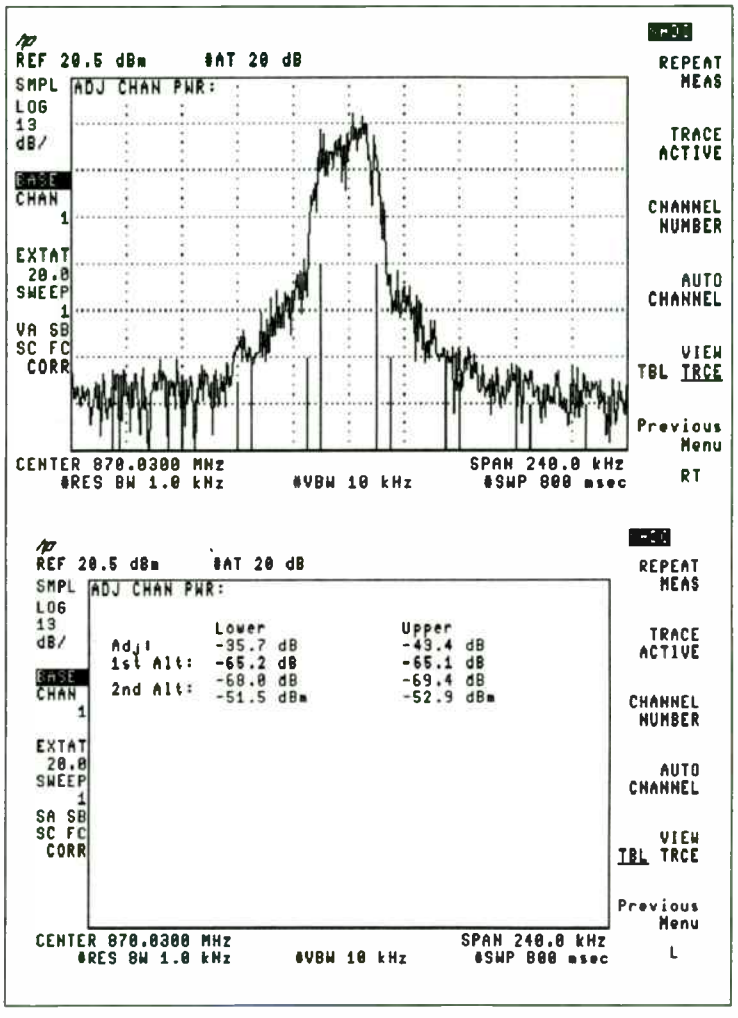


Figure 17: ACP Continuous Carrier

The top HP 8593E spectrum analyzer screen shows the spectrum of an NADC transmitter at 870.03 MHz. The carrier channel and adjacent channels are marked by vertical lines at the lower edge of the screen. The bottom spectrum analyzer screen shows the table of adjacent channel powers resulting from the NADC-TDMA personality ACP test.

Measuring adjacent channel power on burst TDMA signals is more complex than in the continuous carrier case. Burst amplitude modulated carriers produce transient spectra whose power adds to that of the digital modulation. These transient spectra often obscure the lower level digital modulation spectra, making burst adjacent channel power tests difficult. Abrupt power transitions create more transient spectra than slower, more gentle burst ramps. Digital TDMA communication systems usually require slow burst ramps to avoid increasing the adjacent channel power due to transient spectral splatter. An unusually high adjacent channel power may indicate a burst shaping problem.

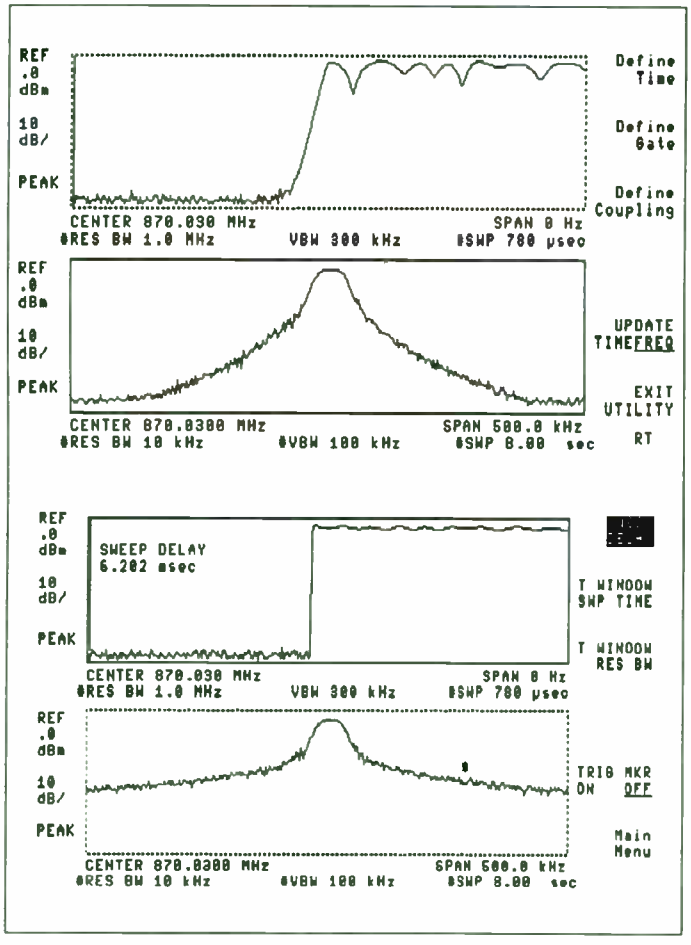


Figure 18: Transient Spectra

The pictures in Figure 18 show transient levels detected in a spectrum analyzer from a slow burst ramp and from an abrupt ramp. Note that a large portion of the spectral splatter displayed is formed within the spectrum analyzer itself.

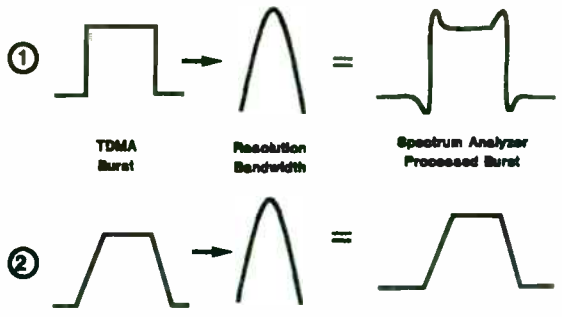


Figure 19: TDMA Burst Interaction with Spectrum Analyzer Resolution Bandwidth

Transient spectra may be formed within the spectrum analyzer. As the TDMA burst passes through the resolution bandwidth filter during a spectrum analyzer sweep, the pulse shape will become distorted at the edges, particularly at frequencies near the beginning and the end of the spectrum analyzer sweep. This distortion is a natural consequence of the way a traditional spectrum analyzer measures the spectrum. As the spectrum analyzer steps through the frequencies in the sweep, the burst is convolved through the impulse response of the filter at each sample point. The convolution of the burst with an off-center-tuned resolution bandwidth filter produces peaks at the edges of the pulse. The distortion peaks, or "ears," are recorded by the analyzer peak detector, so a trace sample point reflects the peak power of the "ears" rather than the true stable power of the burst. See example 1 in Figure 19. A burst with a slow ramp will suffer less distortion, yielding less peaking of the power readings of each sample. See example 2 in Figure 19.

Separating the adjacent channel power due to stable, digital modulation and the ACP due to transient spectra induced by the pulse in the spectrum analyzer is essential to correctly assessing the interference due to spectral spillover in adjacent channels. Note that the spectrum analyzer distortion peaks are positioned near the edges of the burst. A true reading of the power level of the burst is possible away from the edges of the burst.

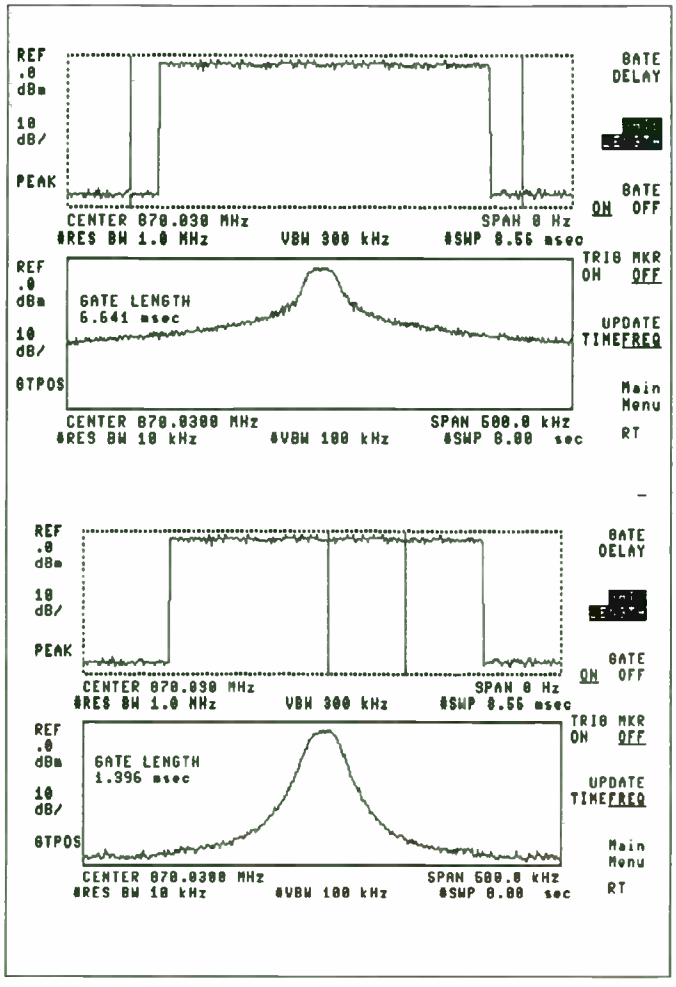


Figure 20: Time Gate Focus on a Stable Portion of the Burst

In the HP 8590 series spectrum analyzers, card cage Option 105 permits the analyzer measurement to be time-limited, or gated, to include only the amplitude-stable portion of the burst. Every point in the trace reflects samples only from the undistorted part of the burst, removing the undesired effects of spectrum analyzer measurement.

The Option 105 time gate adds a switch in the signal processing path of the analyzer. When the switch is closed, the signal passes through to the peak detector and sampler. When the switch is open, no signal is passed to the sampler. The time position of the switch closure, or gate, is easily adjusted from the front panel, using a gate positioning utility.

The spectrum analyzer screens shown in Figure 20 display both the time domain and the frequency domain effects of the time gate. In the pair of spectrum analyzer screens on the top, above, the time gate was positioned to include samples from the entire burst, including the distortion peaks. Using vertical bars, the time gate position is shown in the upper half-screen in time domain. Since the time gate allows samples of the power in the "ears," the peak detector records a high transient level, including that of the "ears." In the lower left, the frequency domain trace shows these high transient spectra. In the pair of spectrum analyzer screens on the bottom, above, the time gate was positioned to direct samples to the amplitude-stable portion of the burst only. The upper time domain picture in the upper right shows the time gate position near the center of the burst. The lower frequency domain half-screen shows the channel spectrum without the effects of spectrum analyzer transients.

In addition to positioning the time gate, a complete spectrum analyzer setup for a burst adjacent channel power measurement includes a few more optimizations. The sweep time must be long enough to catch at least one burst per trace sample value. This means the sweep time should be set greater than the number of trace values times the burst repetition interval. The HP 8590 instruments have 400 trace values per sweep. The peak detector is used to catch and hold the power of the "on" time of the burst. As in the continuous carrier ACP case, a narrow resolution bandwidth is chosen. Dynamic range is often extended through a downloadable program.

Figure 21 , on the next page, is a burst adjacent channel power measurement on a Pacific Digital Cellular (PDC) signal. (Note that PDC was formerly known as JDC, Japan Digital Cellular) The screen on the top shows two traces: the higher level trace with time gating distortion peaks; the lower level trace using the time gate to remove distortion peaks. The HP 85720A JDC measurements personality displays a table of ACP values calculated from the two traces, separating ACP due to modulation from ACP due to transient spectra.

Note that the correct integration equation for transient spectra is calculated with an impulsive noise power integration equation. A total power is sometimes calculated by adding the ACP due to modulation to the ACP due to transients. The total is a result of adding two peak powers with different time characteristics, and should not be considered an RMS adjacent channel power.

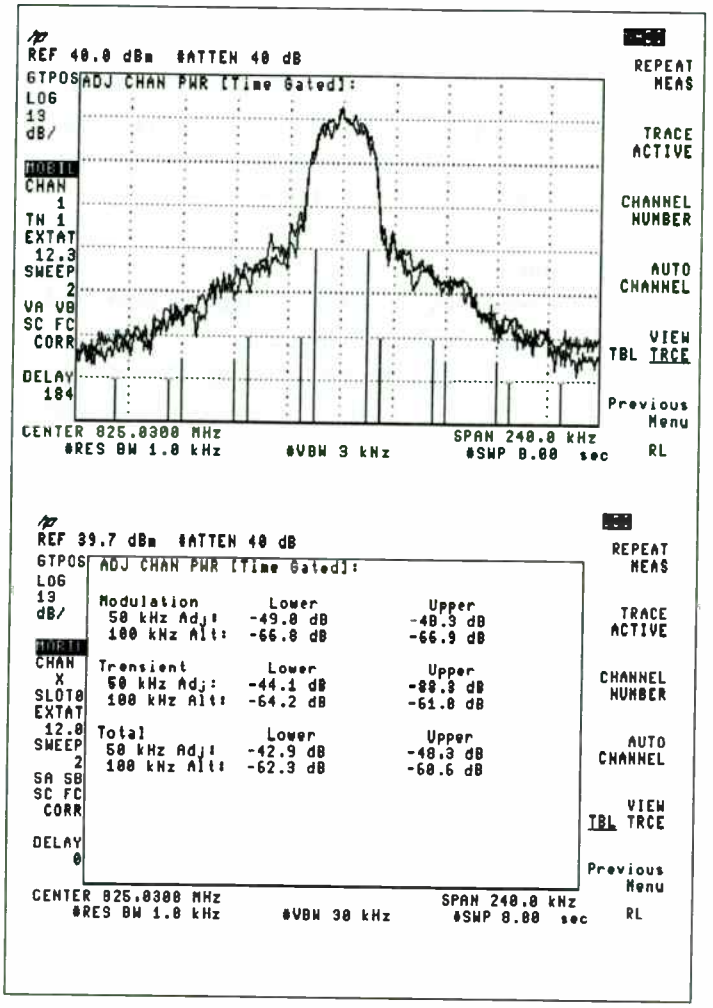


Figure 21: Burst Carrier ACP Test

Both continuous carrier ACP and the intricate burst carrier ACP measurements can be obtained by simple option additions to the HP 8590 series spectrum analyzers. In the card cage, fast sweep and time gate enable ACP in hardware. Specialized ACP tests can be obtained by loading the HP 85718A NADC or HP 85720A JDC personalities. A general adjacent channel power is available as a standard measurement from the spectrum analyzer front panel.

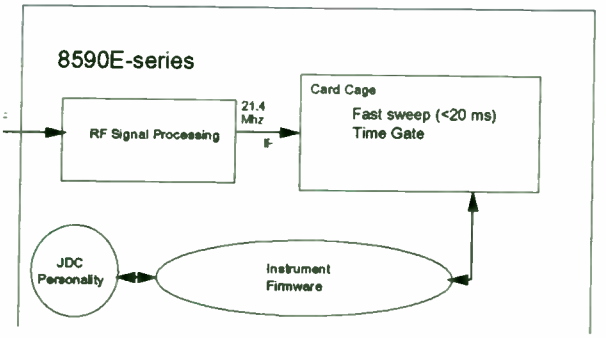


Figure 22: ACP Test Solution

Spurious interference far away from the digital RF transmission frequencies is as important to system performance as adjacent channel interference. Spur searching is a traditional spectrum analyzer measurement, one which the instrument is optimized to perform automatically.

However, there is an important point that is often overlooked. The normal spectrum analyzer auto-sweep time will ensure that continuous spurious signals have the correct amplitude, but burst spurious may be underrepresented or even missed. A spur search for burst spurious signals requires a slower sweep time as given by the equation above. For example, NADC and PDC have a pulse repetition interval of 20 ms. For a resolution bandwidth of 1 MHz, the sweep time must be at least 20 s

To avoid missing spurs during analyzer sweep:
Set spectrum analyzer:

$$\text{Sweep time} \geq \text{PRI} \times \frac{\text{Span}}{\text{Resolution Bandwidth}}$$

for a 1 GHz sweep.

where:
sweep time = spectrum analyzer sweep time (sec)
span = the frequency range of the analyzer sweep
resolution bandwidth = the measurement bandwidth of the analyzer (Hz)
pulse repetition interval (PRI) = time span between bursts (sec)

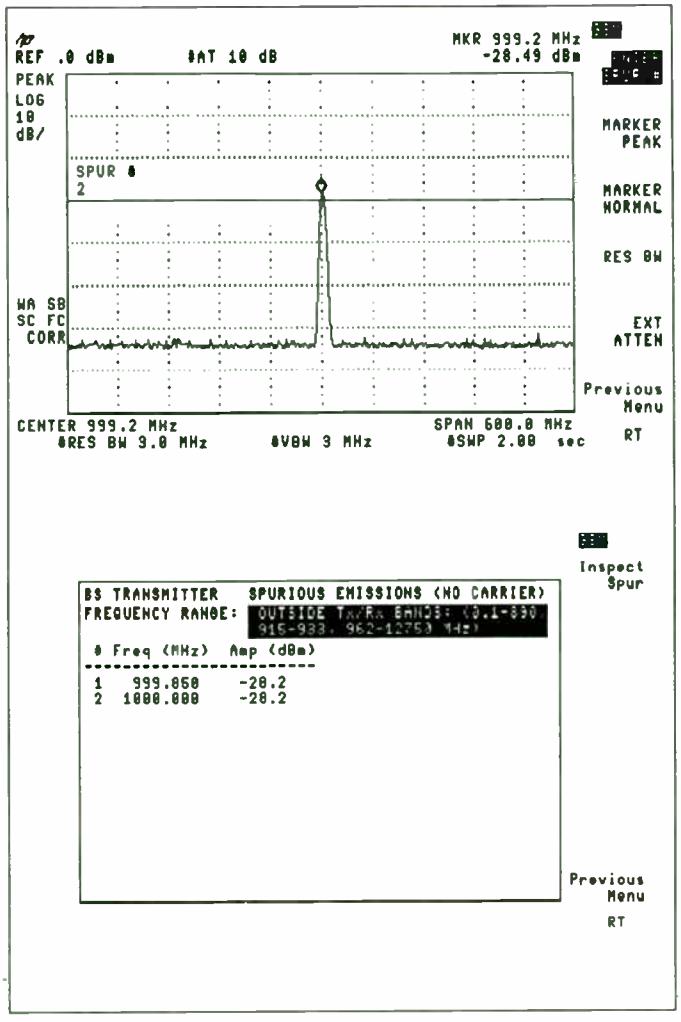


Figure 23: GSM Spurious Emissions Test

Figure 23 shows a spurious emission test performed on the HP 8590 series spectrum analyzer as defined by the GSM standard documents 11.10 and 11.20. The HP 85715A GSM measurement personality performs the complete spur search and highlights in a table any spurs that fail the GSM spec. The user may then examine each spur in the table using a spur inspection utility in the DLP, as shown in the picture on the right.

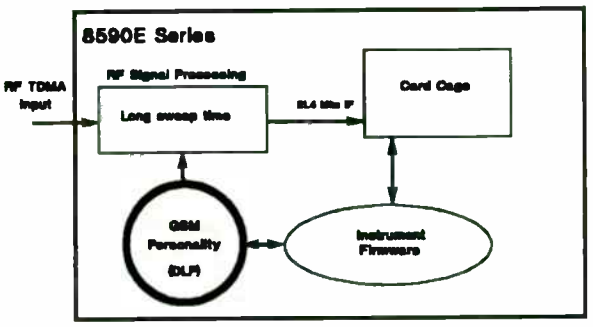


Figure 24: Spurious Emissions Test Solution

The GSM personality is needed to perform a spurious test in the spectrum analyzer. Tailored spur searches for other formats are also available.

Assessing the Quality of Digital Modulation

The last area of digital TDMA communication system test we will examine is assessment of the quality of digital modulation. High quality voice and data transmission rely on accurate digital modulation.

A few simple analog-like modulation quality metrics can be determined independent of the exact digital modulation structure. Two of these are peak frequency deviation and mean frequency error, metrics used in the CT-2 cordless telephone standard. CT-2 employs approximately Gaussian filtered binary phase shift keying (BPSK) to transmit 72 Kbits per second.

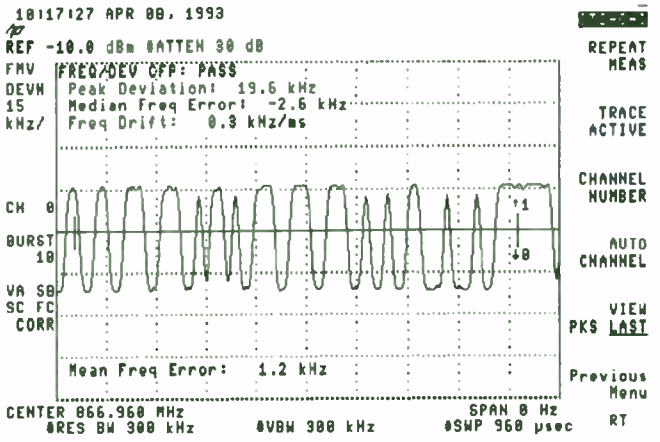


Figure 25: CT-2 Frequency Deviation

The peak frequency deviation in the CT-2 signal is measured using a simple FM detector. The screen in Figure 25 shows the demodulated FM on the CT-2 carrier. The HP 85717A CT-2 personality reads out the peak frequency deviation on screen and calculates the median frequency error.

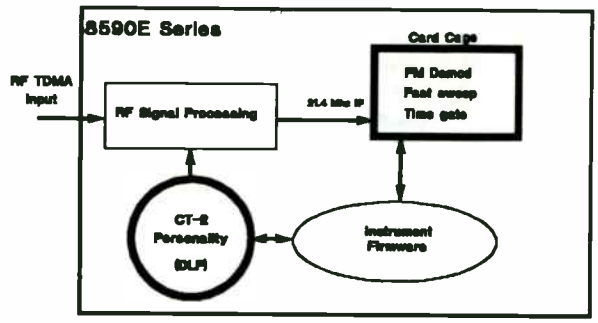


Figure 26: CT-2 Modulation Metrics

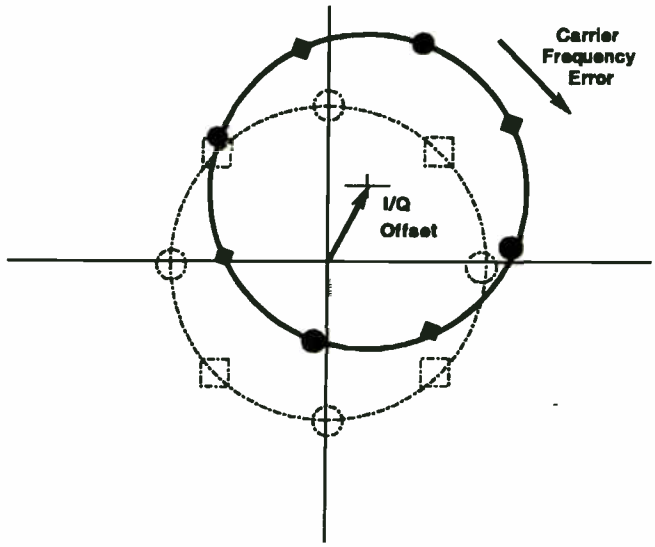


Figure 27: NADC Modulation Metrics

With the addition of the FM demodulator, a fast sweep option, and the time gate, the CT-2 spectrum analyzer system hardware is complete. The CT-2 personality provides convenient automatic tests and compares results to CT-2/CAI standard limits.

Most digital modulation metrics do depend on the modulation format, and sometimes even on the exact bit sequence being sent. The two most common digital modulation forms for the larger communications systems are pi/4 DQPSK and GMSK. Pi/4 DQPSK is used in the NADC, PDC, and PHP standards. GMSK is the modulation for GSM, DECT, and in a simplified form in CT-2. For pi/4 DQPSK, the key quality metrics are RMS and peak error vector magnitude (EVM), carrier frequency error, and I/Q origin offset. For GMSK, the measures include global phase error and carrier frequency error. In each case, we would like to measure these metrics on any time slot, directly from the RF carrier.

The remaining portion of this section will concentrate on measuring the pi/4 DQPSK form of digital modulation in the NADC-TDMA cellular communication system. Figure 27 shows the decision states of an NADC modulation constellation. A perfectly modulated NADC carrier has a baseband I/Q constellation with eight states, divided into two groups. At each decision point the differential phase and the amplitude of the modulation lands exactly on one of the eight states, after compensation with a root-Nyquist filter. A decision point is the point in time at which the constellation state, or symbol, can be read. An ideal constellation is shown with dotted lines above.

The solid lines show a decision point in a transmission with some I/Q offset and frequency error. A real transmission may have a baseband constellation with some inphase (I) and quadrature (Q) channel imbalance, or DC component. The constellation will be shifted away from the origin, where I and Q are zero. This is I/Q offset. Frequency error is exhibited by a slow rotation of the eight transmitted states about the constellation center. This represents a linear phase gain as time increases.

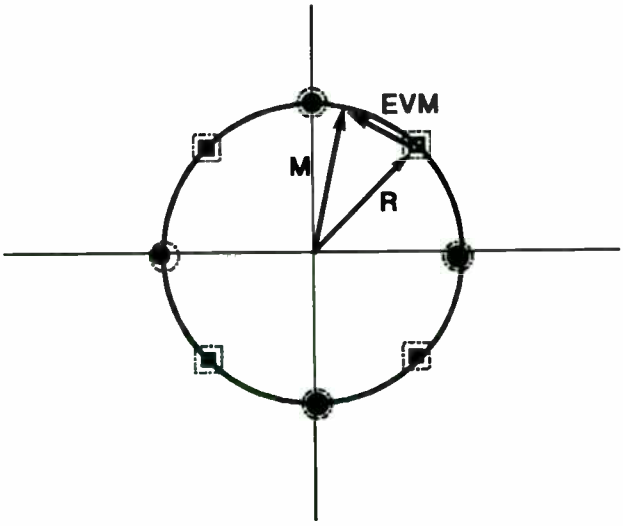


Figure 28: NADC EVM

Error vector magnitude (EVM) is the distance from the ideal state to the measured transmitted state after I/Q offset and frequency error are removed. Amplitude droop, or rolloff, in a burst transmission is also factored out of the EVM. EVM is usually calculated as an RMS value across the symbols in the time slot. The peak EVM achieved at an individual time slot may also be important. A large EVM will cause a symbol to be incorrectly detected, introducing bit errors into the data stream. Voice quality will be degraded.

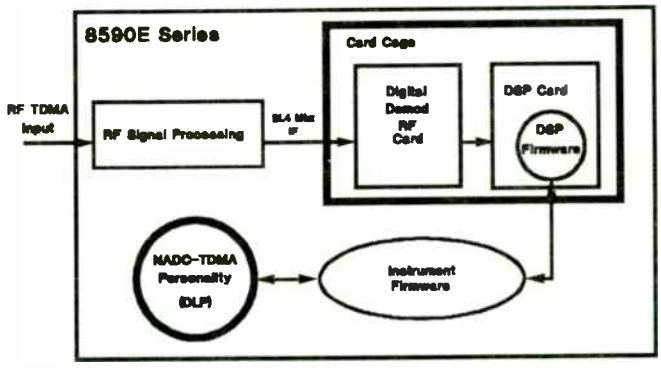


Figure 29: 8590E Digital RF Demodulation System

A new digital RF demodulation capability (digital demod) can be added to the HP 8590 E-series spectrum analyzers for measuring NADC quality. This feature uses digital signal processing to demodulate the digital demodulation. The hardware that must be added to the spectrum analyzer consists of two card cage cards: an RF downconversion card and a digital signal processing card (DSP). The modulated carrier at the input to the spectrum analyzer is downconverted to the spectrum analyzer IF frequency, 21.4 MHz. The RF downconversion card translates the spectrum analyzer IF to 213 kHz for sampling by a fast ADC. The DSP card receives these samples and processes them using a Motorola 56000 DSP chip dedicated to digital modulation measurements.

The digital signal processing algorithms are contained in DSP firmware on the DSP card. This compact firmware is optimized for fast digital signal processing, taking advantage of the parallel processing available on the 56000 DSP chip. The DSP firmware communicates through the spectrum analyzer firmware. Measurements tailored to NADC are driven through a new NADC personality (the HP 85718B), which provides the user interface to the digital demod system. The NADC personality controls the digital demod system, displays the metrics, and compares the results to NADC specification limits.

The HP 8590 E-series NADC digital demodulator system implements a coherent demodulation for NADC pi/4 DQPSK. The downconverted carrier frequency is recovered from the received signal. The reconstructed carrier frequency is used to downconvert the IF samples to baseband. A root-Nyquist filter is applied to corrected baseband samples to compensate for the transmitter root-Nyquist filter. The measurement algorithm complies with NADC-TDMA standards IS-54, IS-55, and IS-56. A minimum error vector magnitude is calculated after removing frequency error, phase offset, amplitude droop, I/Q origin offset, and amplitude scale mismatch.

To begin an NADC digital demod measurement, the spectrum analyzer is optimized to maximize the accuracy of the results. The spectrum analyzer is tuned to a single fixed frequency (zero span) using the most accurate downconversion tuning (counter lock resolution 1 Hz). Several sweeps are taken to stabilize the downconversion tuning. A wide resolution bandwidth is chosen to minimize phase distortion and amplitude ripple. The NADC measurement resolution bandwidth is 1 MHz. Last, the signal level is set to read 2 dB below the top of screen. This signal level maximizes the dynamic range available to the fixed point 56000 DSP processor for accurate calculation.

The signal is then sampled at 1.458 MHz, 60 times the NADC symbol rate of 24.3 KHz. The sample record sent to the digital demod DSP card can be up to half a frame in length, or 3.33 ms.

Digital signal processing of the sample record begins with synchronizing the time of the sample record to the desired time slot. Next, a sample record focused on the desired time slot is taken, sampling synchronously with the symbol rate of 24.3 kHz. The sample record is processed to detect the bit sequence transmitted in the time slot. Provided the detected bit sequence has no errors, a perfect reference signal can be generated. The reference signal is the ideal baseband modulation trajectory for the exact bit sequence that was transmitted. The reference signal is time matched with the decision points of the measured samples. The reference signal is subtracted from the measured samples. The difference is the residual error in the real transmitted signal. The residual error is processed to separate magnitude and phase error components, such as carrier frequency error and error vector magnitude.

Note that a complete error vector magnitude measurement may be made off-the-air from the transmitted carrier. The measurement is automatically focused on the desired time slot.

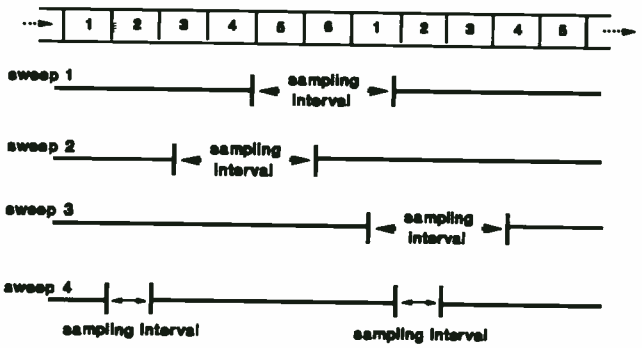


Figure 30: Synchronizing the Measurement

Let us examine the measurement process in more detail, starting with the synchronization of the measurement interval to a user-specified time slot.

A sampling trigger signal is generated on the digital demod cards at the NADC frame rate, 40 ms. The sampling trigger is sometimes called a frame trigger. This trigger frequency is locked to the 10 MHz frequency reference of the spectrum analyzer. The sampling trigger rate is not exactly synchronous to the transmitted frame rate. If, however, the transmitted frame rate is close to nominal, the rate of drift of the sample window relative to the stream of time slots is very slow.

Initially, the NADC digital demod trigger places the sample window at a random offset relative to the desired timeslot, as in sweep 1, 2, and 3 pictured in Figure 30. Remember that sampling intervals occur at the frame rate. To synchronize the sample window, the sample trigger must be offset to place the beginning of the sample window just before the desired time slot. The time slot length is then shortened to the appropriate length. Sampling sweep 4 has been synchronized to time slot 2, focusing the sample window on time slot 2.

The synchronization process requires the digital demod system to find the desired sync word in a stream of transmitted time slots. The sample window is set to capture the longest record of transmitted bits, half a frame or three NADC time slots. A long sample record is taken, and the transmitted bits are detected. The DSP processor then matches a specified sync sequence to the detected bits and determines the time position of the best match. In the NADC system, this requires correlating the 28 bit long sync word with over one thousand detected bits.

Note that if the desired sync sequence is not found in the first half-frame, the sample trigger is repositioned to capture the other half-frame in the sample window. A complete frame is searched to find the sync word.

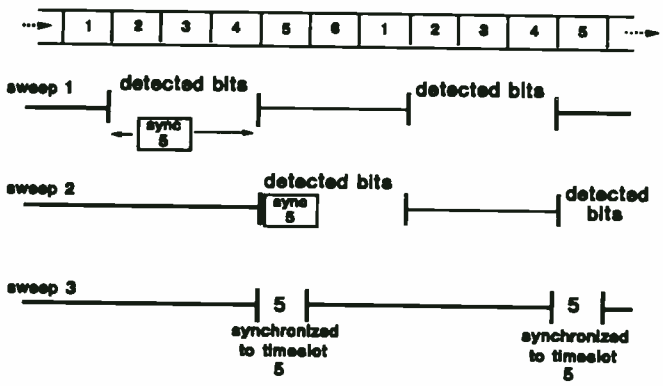


Figure 31: Frame Acquisition

In Figure 31, the digital demod system is synchronizing to NADC sync word five. The sync word is not found in the first half-frame examined. The next half-frame contains the sync sequence near the beginning of the sample record. Knowing the original position of the sampling trigger and the time position of the desired sync sequence in the sample record, the DSP processor can easily determine the time offset to reposition the sample window. The sampling trigger is offset in time from its position in sweep two, placing the sample window at time slot five. The

sample window is shortened for metric measurement on time slot five.

Since the sampling trigger may not be locked to the frame clock of the communication system, the position of the sampling trigger can slowly drift away from time slot five over time. The NADC digital demod system will track this position drift at every measurement. If the drift exceeds a target window, the sampling trigger is automatically repositioned.

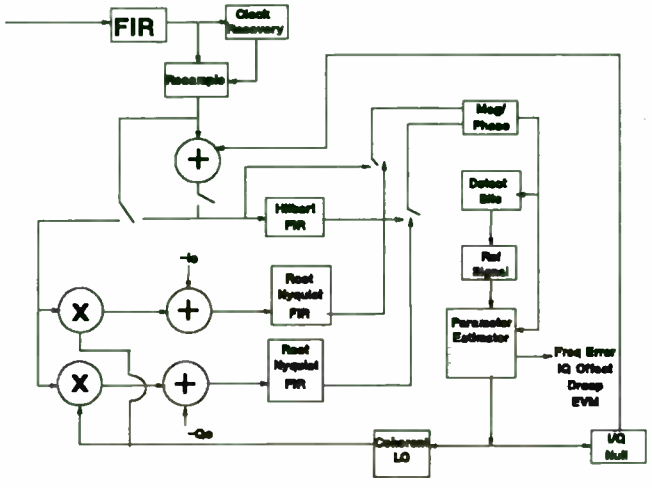


Figure 32: NADC Coherent Demodulation

The synchronized sample record is fed into the NADC coherent demodulation algorithm. A simple block diagram of the algorithm is in Figure 32. The samples are digitally downconverted to baseband in two passes. The first pass uses an initial guess at the correct downconversion frequency. The symbol clock is recovered, and samples are repositioned in time to place one sample per symbol at a decision point. Remember that the decision point is the point in time at which the magnitude and phase of the modulation are at a constellation state where the symbol may be read. In pass 1, the quadrature component is extracted using a Hilbert FIR filter, and sample magnitude and phase are calculated.

Bits are detected from the differential phase of the samples at the decision point. A reference signal is generated. Correction parameters for frequency error, phase offset, amplitude scaling and droop, and I/Q offset are extracted from the difference between the reference signal and the measured samples.

In pass 2, inphase and quadrature coherent local oscillators are generated from the pass 1 corrections to downconvert the measured samples to baseband. The coherent local oscillators are adjusted to remove frequency error, phase offset, amplitude scaling droop, and I/Q offset from the sampled signal. Root-Nyquist filtering is applied to compensate for the transmitter root-Nyquist filter. The magnitude and phase are recalculated. Bits are re-detected and compared to results in pass 1. A new reference signal is created, and residual correction parameters are found. A final correction is applied to the sample record to remove residual traces of frequency error, phase offset, amplitude scaling and droop, and I/Q offset. The RMS and peak EVM are extracted from the final corrected signal and the pass 2 reference signal.

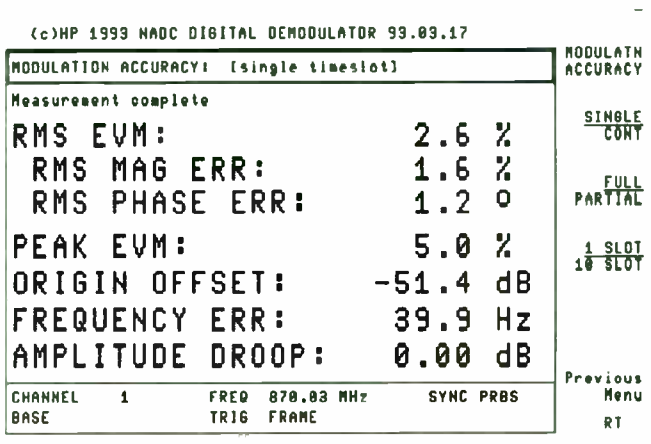


Figure 33: NADC Modulation Metrics

Figure 33 shows some sample output from the NADC digital demodulation system. A summary of metrics is available. RMS error vector magnitude is shown with the magnitude and phase error components. Peak EVM in the time slot is also shown. Carrier frequency error, I/Q offset, and amplitude droop are displayed. Although the initial setup time for this measurement is about 15 seconds, the update rate can be 0.5 to 2 seconds, depending on the exact measurement mode selected. This permits fast feedback on the performance of the NADC-TDMA system.

A baseband modulation pattern diagram displays the modulation trajectory as it travels between symbol states. An example of a pattern diagram is in Figure 34. This is the fully corrected trajectory, reflecting only EVM errors. The corresponding EVM value is shown, so that the user can correlate EVM performance with the appearance of the constellation and trajectory pattern. Gross EVM errors at a single symbol are seen in the pattern diagram. A constellation diagram, showing only the samples at decision points, is also available.

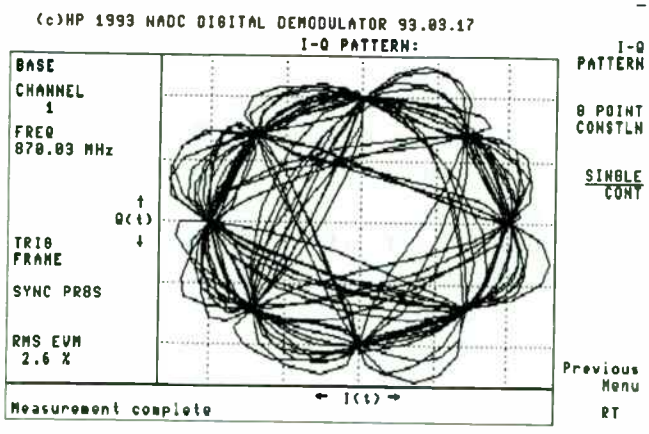


Figure 34: NADC I/Q Pattern Diagram

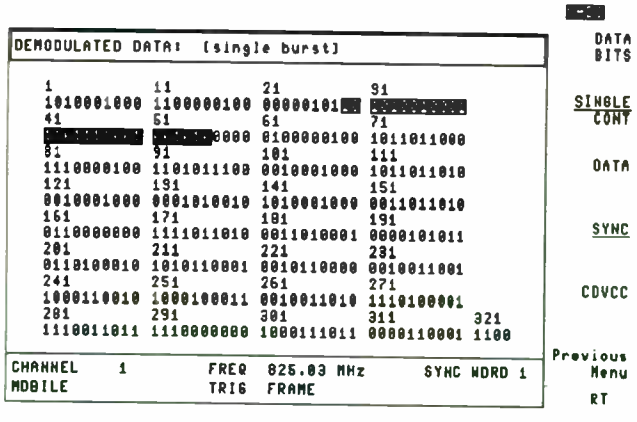


Figure 35: Detected Bits

The detected bits of the time slot can be displayed. Transmitted data, the sync word, or the color code may be highlighted. This screen shows a burst sent from a NADC mobile phone in test mode in time slot 1. Sync sequence 1 is highlighted.

	Digital Demod System Specs	NADC System Performance Limits
EVM Accuracy:		
Max EVM Floor	+1.7 %	12.5 %
Typ EVM Floor	± 1.3 %	
Freq Accuracy:	± 18 Hz + freq ref accuracy x carrier ± 40 Hz typ	± 200 Hz offset (mobile) .25 ppm base station
I/Q Origin Offset Accuracy	.5 dB origin offset for offsets > -40 dB	-20 dBc max (mobile)

Figure 36: Digital Demod System Specs

Compare the accuracy of the HP 8590 E-series NADC digital demod measurements to NADC system limits. An EVM accuracy is specified in two parts: a maximum EVM floor and a positive/negative measurement repeatability. The maximum EVM floor specifies the smallest RMS EVM that the spectrum analyzer can measure. The maximum EVM floor also defines the maximum positive offset added to the true RMS EVM of the signal by the spectrum analyzer. The measurement repeatability specifies the range of displayed values that are possible while measuring a stable EVM source. For, example, if the on-screen RMS EVM reading is 6%, the actual RMS EVM may be as low as 4.3% and not larger than 6%, due to the maximum EVM floor of 1.7 %. The next measurement could have an on-screen reading as high as 7.5% or as low as 4.5%, due to the measurement repeatability of +-1.5`%. If measurements are averaged, the measurement repeatability will average to zero.

The frequency accuracy specification separates the frequency error due to spectrum analyzer downconversion and the digital signal processing limitations from the error due to the accuracy of the frequency reference. This allows users to ascertain the frequency error with an external reference. The basic frequency accuracy of the digital demod is +-18 Hz, plus the frequency reference error at the carrier frequency. For example, PDC users may want to use a rubidium frequency standard to improve the total frequency error accuracy. With the option 004 HP 8590 high-stability reference, the typical frequency accuracy will be +-40 Hz, immediately after calibration of the frequency reference.

The I/Q origin offset can be measured to an accuracy of 0.5 dB down to a floor of -40 dBc.

EVM measurement error comes from both hardware and firmware limitations. The digital signal processing algorithms have an accuracy defined by the filters and algorithms used and by the limits of the fixed point 56000 DSP processor. For example, the ripple of the FIR filters contributes to error in calculating the magnitude component of the EVM. The spectrum analyzer also has some amplitude ripple across the NADC signal bandwidth. The spectrum analyzer frequency accuracy adds to phase and frequency measurement errors. Carrier frequency error accuracy is dominated by the accuracy of the frequency standard. Spectrum analyzer frequency stability is the main limitation of the EVM measurement accuracy. The phase noise of the downconversion adds directly to the EVM measured. The maximum EVM floor value is derived from the RMS phase noise level. The measurement repeatability is derived from the standard deviation of the phase noise distribution.

Measurement Personalities	Hardware Options
85715A GSM Measurements	004 Precision Freq Reference
85717A CT2-CAI Measurements	010 Tracking Generator
85718A NADC-TDMA Measurements	101 Fast Time Domain Sweep
85720A JDC Measurements	105 Time Gate
85718B NADC with Digital Demod Measurements	110 CT2 Demodulator
85713B Digital Radio Measurements	151 + 161 Digital Demod for NADC
	050 Improved Amplitude Accuracy for NADC/JDC

Figure 37: Options Used in Examples

Conclusions

We have seen that the spectrum analyzer provides a platform for time domain, frequency domain, and modulation domain testing of digital RF communications systems. The analyzer may easily be configured to accommodate solutions tailored to many current formats, including GSM, DCS-1800, CT-2, DECT, NADC, and PDC. The spectrum analyzer platform also accommodates future upgrades to expand capabilities to new formats.

Figure 37 is a summary list of the hardware and software solutions that have been presented.

References

1. R. Birgenheier, "Measuring the Modulation Accuracy of pi/4 DQPSK Signals for Digital Cellular Transmitters," Hewlett-Packard Journal, April 1991, pp 73-82.

2. H. Chen, "Time-Selective Spectrum Analysis (TSSA): A powerful tool for analyzing TDMA communication signals," Hewlett-Packard RF & Microwave Measurement Symposium, Signal Analysis Division, 1992.

3. EIA/TIA IS-54, "Cellular System Dual Mode Mobile Station-Base Station Compatibility Standard," Telecommunication Industry Association, April 1992.

4. EIA/TIA IS-55, "Recommended Minimum Performance Standards for 800 MHz Dual Mode Mobile Stations," Telecommunications Industry Association, December 1991.

5. EIA/TIA IS-56 "Recommended Minimum Performance Standards fo 800 Mhz Base Stations Supporting Dual Mode Mobile Stations," Telecommunication Industry Association, March 1991.

6. J. Gorin, "Make Adjacent-Channel Power Measurements," Microwaves and RF, May 1992, pp 137-143.

7. D. Hoover, "An Instrument for Testing North American Digital Cellular Radios," Hewlett-Packard Journal, April 1991, pp 65-72.

8. I-ETS 300 131 "Radio Equipment and Systems (RES); Common air interface specification to be used for interworking between cordless telephone apparatus in the frequency band 864.1 MHz to 868.1 MHz, including public access services." European Telecommunications Standard Institute, France, April 1991.

9. MPT 1375 "Common Air Interface Specification," Department of Trade and Industry, London, May 1989.

10. L. Nutting, "Cellular and PCS TDMA Transmitter Testing with a Spectrum Analyzer," Hewlett Packard RF and Microwave Measurement Symposium, Signal Analysis Division, 1992.

11. RCR STD-27A "Japan Digital Cellular Telecommunication System Standard," Research and Development Center for Radio Systems, Japan, January 1992.

Practical Applications of a Low Cost Low Noise GaAs PHEMT MMIC for Commercial Markets

by
A. Ward
Hewlett-Packard CMCD Applications Department
Richardson, TX.
and
H. Morkner
Hewlett-Packard CMCD R&D Division
Newark, Ca.

ABSTRACT

A high performance low cost surface mount ceramic packaged GaAs MMIC for the 1.5 to 8 GHz frequency range is described. The MMIC uses .15 u gate PHEMT devices, self biasing current sources, source follower interstage, resistive feedback and internal impedance matching to produce a state-of-the-art amplifier. Using a minimum of external components, the MMIC provides a nominal 20 dB gain from 1.5 to 8 GHz, a 2 dB noise figure and a nominal power output of +5 dBm from a single dc power supply. With a simple series inductor, the noise figure can be lowered to 1.6 dB over a smaller bandwidth. Working circuits for various commercial applications are presented along with actual test results.

INTRODUCTION

A GaAs monolithic microwave integrated circuit low noise amplifier has been designed for use in the numerous commercial applications in the 1.5 to 8 GHz frequency range[1]. Typical markets include GPS at 1.5 GHz, PCN at 1.9 GHz, MMDS at 2.1 GHz, ITFS at 2.5 GHz, ISM at 2.4 GHz and 5.8 GHz, TVRO at 4 GHz and instrumentation in the 1.5 to 8 GHz frequency range.

The MMIC includes internal impedance matching and integrated biasing allowing simplified low noise amplifier design eliminating the numerous components normally required for a discrete low noise amplifier design. The MMIC delivers an Fmin within a dB of a typical discrete design in a fraction of the space. The MMIC is available in a low cost ceramic package and is now available as the Hewlett-Packard MGA-86576.

The performance of the MMIC as a low noise amplifier will be described. Other special applications such as a variable gain amplifier and active mixer will also be addressed.

DESIGN

The schematic diagram shown in Figure 1 is a lumped element representation of the MMIC. The MMIC consists of two gain stages, an interstage source follower stage, three current sources, two resistive feedback networks, several bias resistors and bypass capacitors. The current sources insure that each PHEMT stage is biased at 25% of Idss. Process variations effect both

the FET and current source simultaneously resulting in a constant device Idss producing devices with very repeatable RF performance. .

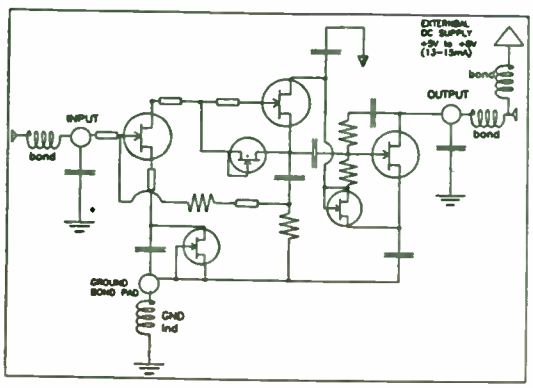


Figure 1. MMIC Schematic Diagram

An advantage of the PHEMT process is the inherent low knee voltages which allow the FET stages to be connected in series to share the device current. This results in a low current MMIC that operates from a supply voltage of only 5 volts.

PERFORMANCE: MMIC VS FET

The MMIC offers several advantages over a discrete FET design. The primary advantage of the MMIC is in the bias configuration. Whereas a typical discrete FET amplifier design may use active biasing to set the bias point, the MMIC requires only a simple RF choke to feed in bias from a 5 volt power supply. This minimizes the board space required to bias the device by eliminating 10 to 15 components. The internal current sources in the MMIC keep the device current bias in a fairly narrow window which results in very repeatable RF performance.

A second advantage of the MMIC is that the input is partially matched to 50Ω making the MMIC easier to match than the generally high impedances associated with the unmatched FET. As an example at 4 GHz, the S11 of the MMIC is less than 0.5 while Gamma Opt is less than 0.4. Typically, the S11 and Gamma Opt of the unmatched FET are much higher at 4 GHz. The lower input impedance generally means it is easier to achieve a wide bandwidth with a low tolerance matching structure. Even with no input matching, the 50 Ω noise figure is about 2 dB. With a small amount of inductance in the input network, it is possible to achieve device noise figure as low as 1.6 dB.

A third advantage of the MMIC is that its gain is equivalent to a two stage FET amplifier. This has additional advantages in the form of decreased component count and decreased board space required.

The disadvantage is its inherently higher noise figure. The MMIC noise figure is generally about 1 dB higher than the typical low noise PHEMT device. This is due to the resistive loading and feedback inherent to the design. For most applications, this may not be a problem. The nominal 2 dB noise figure does make the MMIC an ideal first stage for most applications and an ideal second stage device in a very low noise amplifier.

APPLICATION

BIAS DECOUPLING NETWORKS

Biasing the device simply requires the use of an RF choke to supply 5 volts to the output terminal. The RF choke at microwave frequencies can be in the form of a high impedance microstripline properly bypassed at the power supply end. The optimum length would be a quarterwave at the desired frequency of operation but as the results show one nominal length does provide good operation from 2 to at least 6 GHz. The use of lumped inductors is not desired since tney do tend to radiate and cause undesired feedback.

Low loss capacitors are used to couple the RF in and out of the device. The dc blocking capacitors should provide a low impedance over the desired operating bandwidth without adding excessive series inductance. Although there is no voltage present at the input of the device, it is suggested that a dc blocking capacitor be used especially if the device is preceded with another amplifier device. A suggested value of blocking capacitor for use in the 1.5 to 6 GHz frequency range is 27 pF.

NOISE MATCH

The device's 50Ω noise figure is approximately 2 dB. Decreasing the noise figure at 4 GHz, as an example, requires that a matching network transform 50Ω to Gamma Opt of the device. (Gamma Opt is the device reflection coefficient required for the device to produce its minimum noise figure) At 4 GHz, the Gamma Opt is .38 @ 51 degrees. A Smith Chart exercise suggests a series inductance of about 2 nH and a shunt capacitance less than 0.2 pF to transform 50Ω to Gamma Opt. The best way to implement this matching circuit would be to raise the input lead of the device and make it into a loop . The shunt capacitance required is so low that it can be overlooked and still achieve good results. The matching network will lower the noise figure about 0.5 dB at 4 GHz.

PRINTED CIRCUIT BOARD MATERIALS

Most commercial applications dictate the need to use inexpensive epoxy glass materials such as FR-4 or G-10. Unfortunately the losses of the this type of material can become excessive above 2 GHz. A 0.5 inch long 50 Ω microstripline along with a 27 pF blocking capacitor and two SMA end launch connectors has a measured 0.35 dB loss at 4 GHz and 1.25 dB loss at 6 GHz! We can generally tolerate the loss as a gain loss but not as an increase in device noise figure.

A second concern would be the thickness of the material. In a typical microstripline topology, the common leads of the MMIC must be attached to the bottom ground plane with the use of plated through holes. The inductance associated with these plated through holes adds series inductance which can cause gain peaking and potential instability. The MMIC has been designed to accommodate dielectric board thicknesses of 0.040 inch or less without adversely effecting MMIC operation.

ACTUAL CIRCUITS AND MEASURED PERFORMANCE

GAIN AND NOISE FIGURE PERFORMANCE

The schematic diagram of the demonstration amplifier is shown in Figure 2. The amplifier consists of 50 Ω microstripline and dc block capacitors and a bias decoupling line. The resistor in series with the bias decoupling line can be adjusted depending on the available supply voltage. It is suggested that a minimum of 10Ω be used to de-Q the bias decoupling line.

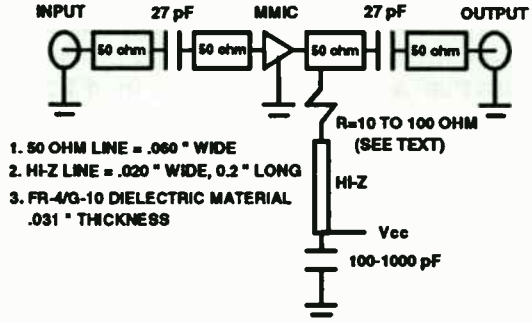


Figure 2 Schematic Diagram of MMIC demonstration amplifier

The gain performance of the MMIC as measured in a demonstration amplifier built using G-10 dielectric material is shown in Figure 3. The gain is shown for device voltages of 5, 6, 7, and 10 volts.

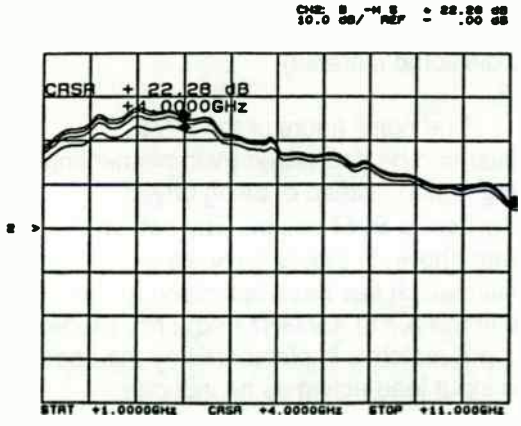


Figure 3. Gain vs. frequency for G-10 demonstration amplifier

The extraordinary band width of this amplifier makes it usable anywhere in the 1 to 10 GHz frequency range. The gain peaks

between 3 and 4 GHz with about 22 dB gain at a device voltage, Vdd, of 5 volts. Increasing, Vdd, to 7 volts increases the gain at 3 GHz to 26 dB. The noise figure of the device was optimized for lowest noise in the 2 to 6 GHz frequency range. Actual untuned noise figure at 2.3 and 4 GHz including board losses is 2.5 dB. Subtracting the 0.35 dB loss for the input microstripline plus dc blocking capacitor suggests an untuned device noise figure 0f 2.15 dB at 4 GHz.

For applications that require a lower noise figure, the use of a low loss material such as Duroid™ or Taconics TLY-5 is highly recommended. Measured noise figure of a demonstration board using Duroid 5880 is shown in Table I. The amplifier exhibits less than a 2.54 dB noise figure from 1.3 to 5.8 GHz. The low noise performance of this device makes it suitable for low noise amplification in the 2.4 and 5.8 GHz spread spectrum bands. Noise figure at 900 MHz increases to 2.97 dB while noise figure at 10.5 GHz is still a respectable 3.34 dB.

FREQ(GHz)	Noise Figure(dB)
0.9	2.97
1.3	2.46
1.6	2.27
2.3	2.10
4.0	2.04
5.8	2.54
10.4	3.34

Table I Noise Figure vs. frequency (Duroid 5880 dielectric material)

The noise figure of the MMIC amplifier can be decreased by implementing a noise match instead of being driven directly from a 50 Ω source. The schematic diagram shown in Figure 4 shows an amplifier which has been optimized for low noise in the 3.7 to 4.2 GHz frequency range. The input match is implemented by the use of the input lead acting as an inductor.

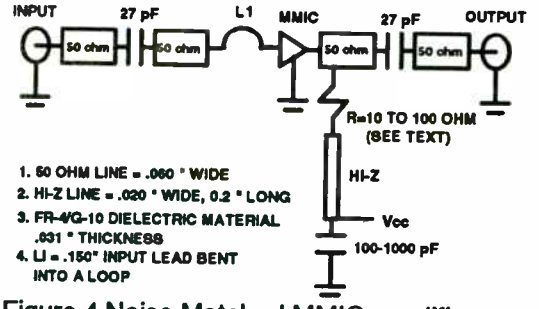


Figure 4 Noise Matched MMIC amplifier (G-10 dielectric material)

The performance of the circuit shown in Figure 4 is shown in Table II.

FREQ(GHz)	GAIN (dB)	N.F.dB)
3.7	24.0	1.92
3.8	24.0	1.92
3.9	23.7	1.89
4.0	24.0	1.89
4.1	23.9	1.90
4.2	23.3	1.86

Table II Gain and Noise Figure for MMIC amplifier with noise match (G-10 dielectric material)

Subtracting the 0.35 dB loss of the G-10 dielectric material suggests an actual noise figure of 1.55 dB for the device. To achieve within a 0.1 dB of this noise figure in an actual board will require the use of one of the lower loss materials discussed earlier.

POWER OUTPUT PERFORMANCE

Although primarily designed for low noise and broad band stable gain, the MGA-86576 provides moderate power output considering its low bias point. At a Vdd, of 6 volts, the device provides a measured output 1 dB gain compression point, P1dB, of +5 dBm at 4 GHz. At a higher Vdd of 7 volts, P1dB increases slightly to +5.5 dBm. The measured two-tone third order intercept point , IP3, as referenced to the output is +16 dBm. Increasing the Vdd to 7 volts increases IP3 to +17 dBm.

OTHER PROPERTIES

Although specifically designed for use as a low noise amplifier, the MMIC also works well as a variable gain amplifier and an active mixer. The following section discusses the MMIC in these two applications and presents actual but not guaranteed performance data.

VARIABLE GAIN AMPLIFIER

An added benefit of the MGA-86576 is its ability to operate as a variable gain amplifier. By simply adding a 7.5 K ohm chip resistor and an additional bias decoupling line appropriately bypassed, an additional positive or negative voltage can be injected into the input terminal of the device for manual gain control. The device can only withstand a small voltage at this port.

Figure 5 shows the decrease in gain in 5 dB steps as the input voltage is increased to a maximum of +.430 volts. Figure 6 shows the decrease in gain in 5 dB steps as the input voltage is increased to -.762 volts. With a slight negative voltage applied to the input terminal of -.093 volts, the gain actually increases about 2.7 dB before decreasing with increased negative voltage. The graphs also indicate that the gain reduction versus frequency is fairly constant over a very wide bandwidth.

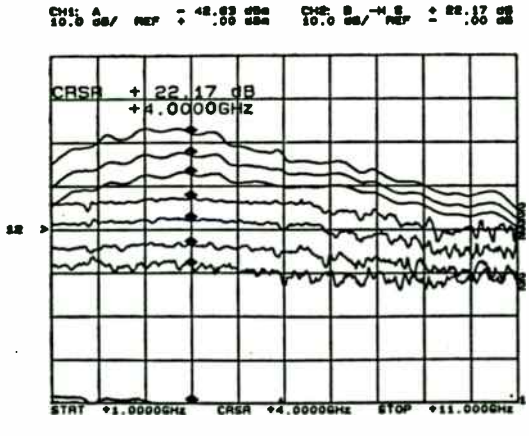


Figure 5 Variable gain operation with positive control voltage

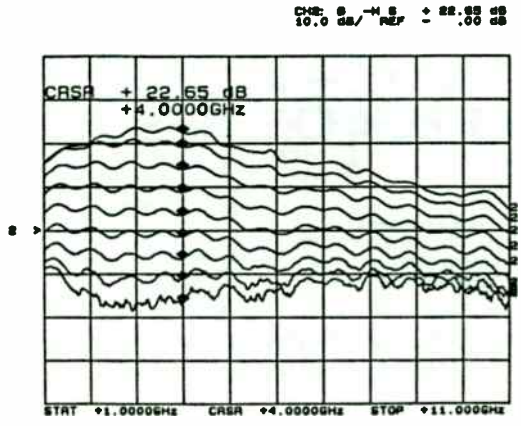


Figure 6. Variable gain operation with negative control voltage

Using a negative control voltage as opposed to a positive voltage for gain control actually is better if lower noise figure at reduced gain is desired. As an example at 4 GHz, when using a negative voltage to decrease the gain by 15 dB, the noise figure increases to 3.9 dB. Using a positive voltage to decrease the gain by 15 dB increases the noise figure to 9.9 dB.

12 GHz ACTIVE MIXER

The MMIC was also tested as an active downconverter for use in DBS applications at 12 GHz. A schematic diagram is shown in Figure 7. The circuit is etched on low cost G-10 dielectric material. The MMIC was configured as a drain or output pumped mixer by injecting a +11 dBm 10.8 GHz local oscillator signal into the output port. The input port is used as the RF port while the IF is coupled out the output port. Quarterwave microstriplines are used at the output port to achieve LO to IF isolation. The mixer achieved a SSB noise figure between 11 and 12 dB and a conversion gain between 1.5 and 2.5 dB over the entire 11750 to 12250 MHz frequency range. The noise figure performance can be improved by several dB by using lower loss material. Considerably better performance is possible at the lower frequencies where dielectric losses are lower and the basic noise figure of the device is a couple of dB lower.

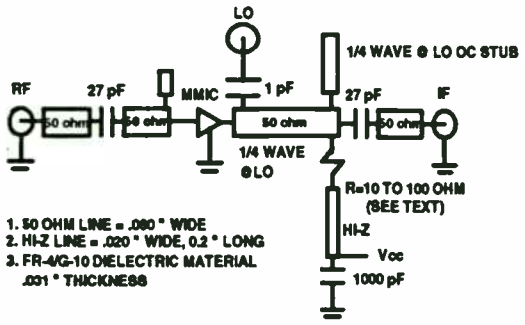


Figure 7.MMIC DBS downconverter (G-10 dielectric material)

The measured LO to IF isolation is 17 dB while the measured LO to RF isolation is 23 dB. The LO to IF isolation is obtained by the use of simple quarterwave microstriplines in the output network while the S12 of the device contributes to the LO to RF isolation of the mixer.

CONCLUSIONS

GaAs PHEMT technology is revolutionizing the implementation of low cost high performance low noise amplifier designs for commercial applications. Surface mount packaged MMICs are achieving noise figures and gains that rival discrete amplifier designs.

REFERENCES

1. H. Morkner, M. Frank, D. Millicker, "A High Performance 1.5 dB Low Noise GaAsPHEMT MMIC Amplifier for low cost 1.5 - 8 GHz commercial applications", 1993 IEEE MONOLITHIC CIRCUITS SYMPOSIUM, PG13

HIGHLY INTEGRATED GaAs MMIC RF FRONT END FOR PCMCIA PCS APPLICATIONS

J. Mondal, S. Geske, T. Kotsch, A. Laundrie, G. Dietz, E. Bogus, J. Blubaugh, S. Moghe

Northrop Corporation
Electronics Systems Division Rolling Meadows Site (ESD-RMS)
600 Hicks Road
Rolling Meadows, IL 60008

Abstract

A fully integrated GaAs Microwave Monolithic Integrated Circuits (MMIC) transceiver chip for use on Personal Computer Memory Card International Association (PCMCIA) card in wireless modem and Local Area Network (LAN) aplications is described. The chip is small in size ($100x150mil^2$) and exhibits good performance after the first fabrication pass. An ongoing activity to further improve the electrical performance, reduce the size of the chip and develop a low cost composite package using multilayer ceramic microwave integrated circuits (MCMIC) is also outlined. A proposed system-specification and -architecture is presented.

1. Introduction

This paper describes the design, fabrication, and performance of a highly integrated GaAs MMIC radio frequency (RF) front-end. Designed specifically for wireless modem and LAN applications, the front-end achieves the performance and small size necessary to contain all RF and digital hardware on a single PCMCIA card. Such cards are increasingly used to expand the capabilities of notebook and palm-size computers.

The RF front-end employs a highly integrated transceiver chip. The GaAs Integrated Circuit (IC) includes an upconverter, a medium-power output amplifier, a transmit/receive switch, a low-noise input amplifier, and a downconverter. A Voltage Controlled Oscillator (VCO) is also included on the PCS transceiver chip to supply an LO signal to both the upconverter and the downconverter. The chip is designed to operate from 800 to 1800 MHz.

The Personal Communication System (PCS) transceiver IC described exhibits good performance after first-pass fabrication. The front-end utilizes the MMIC on a MCMIC substrate. This approach should significantly reduce the size of the front end. MCMIC will include a synthesizer IC for the transceiver LO, as well as biasing and filtering components not included on the transceiver chip.

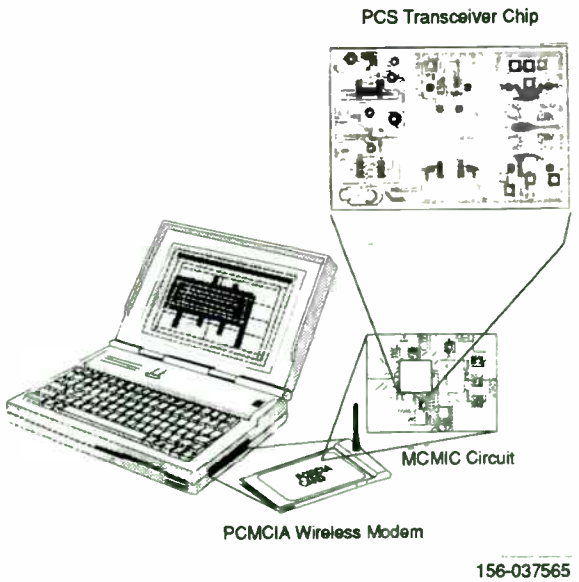


Figure 2-1. Wireless data communications using highly integrated PCS MMIC chip.

2. Applications

The transceiver MMIC demonstrates the highly integrated GaAs MMIC subsystem development capabilities of the Advanced Microwave Technology group at Northrop Electronic Systems Division (Rolling Meadows, IL). The technology under development is suitable for a variety of commercial applications, including PCS and Intelligent Vehicular Highway Systems (IVHS), and also for military Electronic Warfare (EW) systems.

The great demand for portable phones during recent years is expected to continue, as more wireless modem and LAN products are offered to the public. One such wireless data communication device, that is being developed at ESD-RMS, is illustrated in Figure 2-1.

3. GaAs MMIC PCS Chip

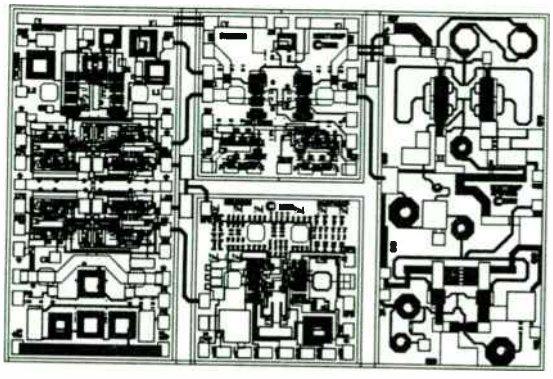


Figure 3-1. Transceiver MMIC with all RF functions for wireless LAN/modem.

The transceiver MMIC (Figure 3-1) is a fully integrated chip on a GaAs-substrate. It measures 100 x 150 mil^2 . The chip is designed to cover the bandwidth 0.8-1.8 GHz. It has all the essential RF functional blocks (Figure 3-2) for a transmit/receive (T/R) RF front end in wireless modem/LAN and cellular applications. In this section the functional blocks will be discussed, and measured data will be presented for individual blocks. The measured data were taken after the first pass fabrication of the chip. The chip is being slightly redesigned to accommodate additional functions and to reduce the size by 33%. The final chip size will be 100 x 100 mil^2. The main advantages of a higher level of integration are smaller size, lower cost, high reliability and ease of assembly.

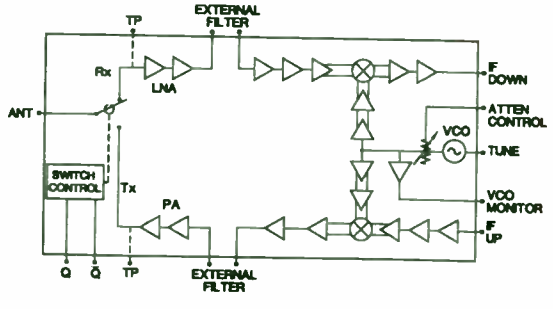


Figure 3-2. Essential RF functional blocks on transceiver chip.

The receive path of the transceiver chip (Figures 3-1 and 3-2) consists of two stages of low noise amplification, followed by an external filter (optional), a differential down converter and two stages of intermediate frequency (IF) amplification. The transmit path has three stages of IF amplification followed by differential up-conversion and two stages of power amplification. The transmit/receive (T/R) switch at the output routes transmit and receive signals to and from the antenna, respectively.

The individual building blocks of the integrated PCS chip have been characterized and the performance data are shown below. The whole PCS chip is being mounted in a package for full charaterization as an integrated chip.

Description and Performance of Individual MMIC in the Transceiver Chip

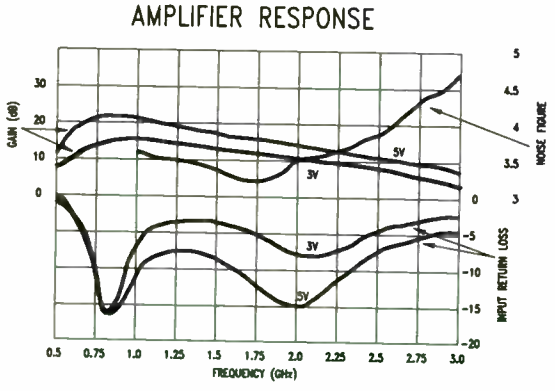


Figure 3-3. Gain, Noise Figure and Return loss of the LNA in the transceiver chip.

Low Noise Amplifier (LNA): The LNA has two stages of amplification and is designed to operate from 0.8 to 2.4 GHz. It operates from a single positive supply (3-5V) with low power dissipation (30-50 mW). The active devices are enhancement metal semiconductor field effect transistors (MESFETs). Resistive biasing is used and it decreases the size of the chip at a slight cost of dissipated power and reduced efficiency. In the receive path, the total current drawn from the supply by the LNA is 10-12 mA. The MESFETs, after the resistive drop in the drain bias circuits operate very close to the knee voltage of the I-V curves. The measured performance of the LNA is shown in Figure 3-3. The performance is shown at both VD=3V and 5V. At VD=5V, the amplifier shows a gain of 20 dB from 0.8 to 1.25 GHz and it rolls off to 12 dB at 2.6 GHz. At VD=3V, the gain is between 10 and 15 dB over most of the 0.8-2.4GHz band.

The noise figure is less than 3.5 dB over this same 0.8-2.4 GHz bandwidth. The chip is redesigned for the second iteration to deliver a gain of 20 dB at VD=5V and a noise figure of 3 dB from 0.8 to 2.4 GHz.

Down-, Up-Converter: The converter MMICs form a part of a bigger section in the transceiver chip that also has the LNA and a 90° splitter. The 90° splitter is not being used on this iteration. The splitter will be added on the second iteration to provide I-Q outputs.

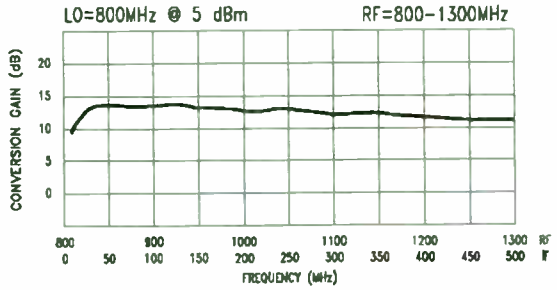


Figure 3-4. Performance of the down-converter in the transceiver chip.

The RF is first amplified by a two stage amplifier and then split by a differential amplifier for feeding to a MESFET-quad mixer. The LO is also split and amplified by a two stage differential amplifier and then fed to the gates of the FET-quad mixer. The IF from the FET-quad mixer is combined and amplified through a differential amplifier to provide single ended output.

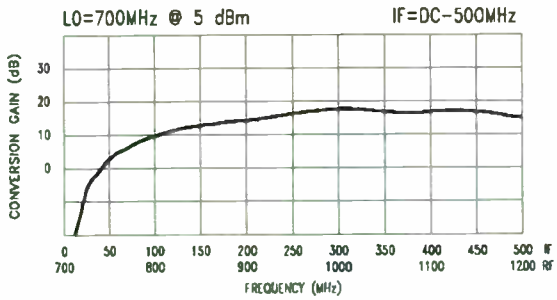


Figure 3-5. Performance of the up-converter in the transceiver chip.

The dc power consumption of the converter is approximately 100 mW for a 5V drain supply. Figure 3-4 shows the down conversion gain with the LO fixed and RF varied. Up to 1300 MHz, the conversion gain is 11-12 dB. Figure 3-5 similarly shows the up-conversion performance. The up-converter is a mirror image of the down-converter along the X-axis. The up-conversion performance degrades at lower IF (less than 100 MHz). Above 100 MHz IF, the up-conversion gain is more than 10 dB.

In the second iteration, the up-conversion gain is being increased to more than 10 dB for an IF down to 30 MHz, by increasing the coupling capacitor between IF amplifier stages.

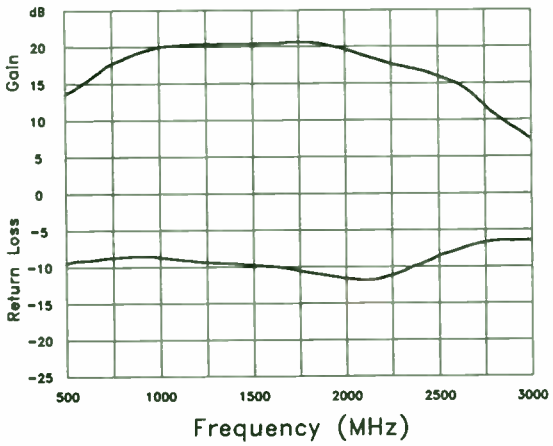


Figure 3-6. Small signal response of the power amplifier in the transceiver chip.

Power Amplifier: The power amplifier in the transmit path is designed to deliver 25-27 dBm of power at 1-3 dB compression points with 18-20 dB of gain over 0.8-1.8 GHz. It uses off-chip coils and capacitors for dc-biasing. The off-chip bias components are necessary to handle the total current (350-450 mA). The on-chip coils have low Qs and dissipate a fair amount of dc-power, thus degrading the efficiency. The MMIC occupies 50 x 100 mil^2. The MMIC does not utilize via holes, the ground is provided by a number of bond wires that connect the ground pads (distributed around the chip) to the carrier plate. The MMIC requires a negative gate supply. A self-biasing scheme can be adopted by connecting a resistor and by-pass capacitor to the ground pad. The self-biasing scheme requires a single bias supply at a cost of degraded efficiency. Figure 3-6 shows the small signal response of the power amplifier. The amplifier shows 17-20 dB gain from 0.8 to 2.0 GHz. Figure 3-7 shows the power output and power added efficiency over frequency. The output power at 1 dB compression is more than 25 dBm from 0.8 to 2.0 GHz. The output power is 25-27 dBm for 1-3 dB compresssion or expansion. The associated power added

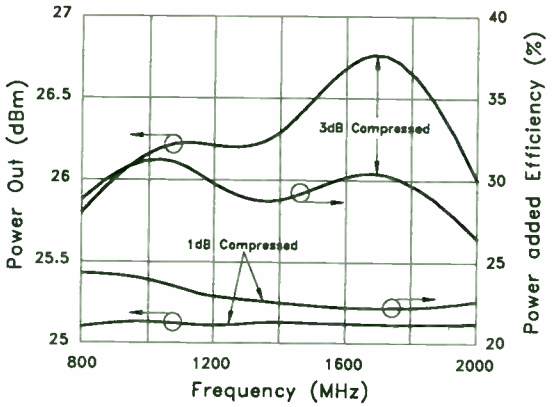


Figure 3-7. Power output and power added efficiency of the power amplifier @P_{1db} and @P_{3db} in the transceiver chip.

efficiencies are between 22 and 30% over the 0.8-2.0 GHz bandwidth. The power output in the second iteration design is being increased to 27 dBm (at 1dB compression point) over the 0.8-2.0 GHz bandwidth.

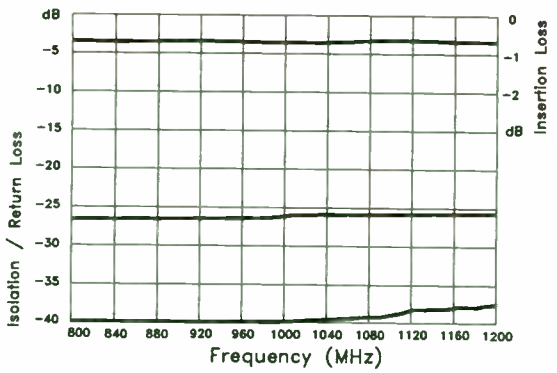


Figure 3-8. Insertion loss, return loss and isolation of T/R switch in the transceiver chip.

Transmit/Receive (T/R) Switch: The T/R switch is a single-pole double-throw (SPDT) type. It routes the signal between the antenna and receiver or transmitter. Its measured performance is shown in Figure 3-8. The switch has an insertion loss of less than 1dB, input output return loss of better than -20dB and an isolation better than 35 dB up to 1.2 GHz. The switch has a 1dB compression point of 24 dBm. In the second iteration, the switch power handling is being increased to 25 dBm and the isolation is being improved to 45 dB.

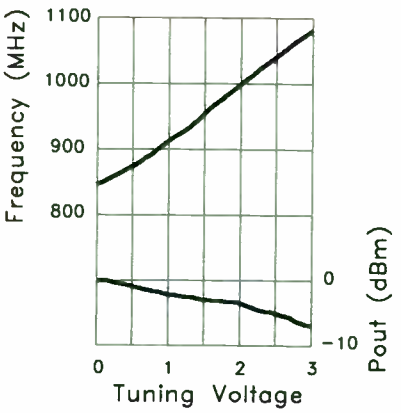


Figure 3-9. Frequency tunability and power output of VCO in the PCS MMIC.

Voltage-Controlled Oscillator (VCO): The on-chip VCO provides the LO for both receive and transmit paths. Tuning is achieved by an off-chip varactor, the capacitance of which is changed by a tuning voltage. The VCO chip also incorporates a buffer amplifier that follows the oscillator. There is an on-chip attenuator (optional) between the oscillator and the buffer amplifier to control the output power. The oscillator and buffer amplifier are designed to draw a total dc current of 18-22 mA from a 3-5V supply. Figure 3-9 presents the measured VCO performance over 0-3V tuning voltage. Total tunability is 840 MHz to 1085 MHz. The power out is 0 to -8 dBm. A redesign of the VCO for the second iteration is underway to increase the power to 6 dBm and tunability from 0.8 to 2.0 GHz range.

Integrated PCS MMIC: The PCS chip is being put in a package as shown in Figure 3-10. The overall system performance as an integrated PCS chip is being evaluated.

4. Example System Architecture

The application for this design is in a PCMCIA card. The product would be housed in a PCMCIA card that plugs into an ExCA compatible port on a portable notebook or handheld computer. There are significant advantages in implementing a wireless data product in a small package. Unrestricted portability is the prime drive of PCMCIA based wireless products. However, along with the smaller package comes the requirement for reduced power consumption. Recent developments in MMIC and MIC

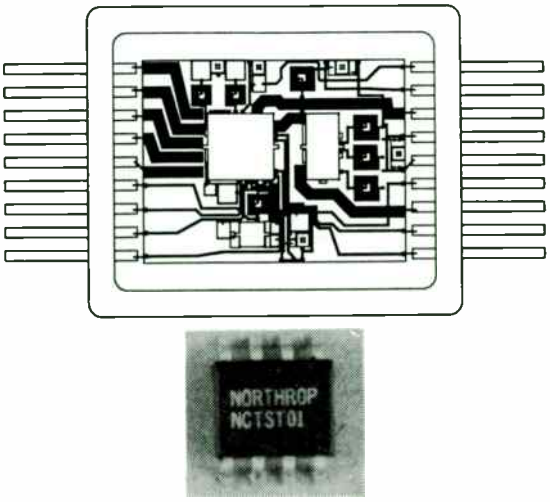


Figure 3-10. The plastic package in which the transceiver MMIC is being inserted along with the multilayer ceramic interfacing circuit.

technologies in the defense industry provide an opportunity to address these requirements.

A proposed modem (Figure 4-1) product would consist of several key functional blocks: 1) PCMCIA interface, 2) digital modem function, 3) spread-spectrum function, 4) RF functional block, and 5) antenna system.

In general, the computer architectures that this wireless modem product will be used with must adhere to the PCMCIA 2.0 standards as well as ExCA interface format compatibility. Most computer manufacturers have adopted these formats as standards for their current and next generation products. The physical form of the module follows along the lines of a Type II extended PCMCIA card. This allows sufficient room to implement the interface, processing, and RF components as well as the antenna system. Due to various host computer physical housing designs, a non-shielded area in the housing for antenna applications cannot be guaranteed. Thus, a small portion of the PCMCIA card may extend out of the host housing.

The system design provides peer-to-peer wireless communications. Such a local wireless network can be used for warehouse inventory-taking, cooperative learning in classrooms, and pen- and notepad-based applications. A peer-to-peer system would consist of two portable computers and two wireless modems. The physical nature of the system is such that its operation would be transparent to the user. Application software and system specific drivers can be easily written to interface with the modem protocols and to provide various specific functions.

PCMCIA Interface: The PCMCIA interface would consist of a 68-pin PCMCIA connector, RAM and ROM, and an interface adapter for communicating card specific functions and configurations to the host computer. The function of the interface follows the card interface structure (CIS) as defined in the PCMCIA 2.0 standard.

Modem Function: The modem function would provide for a maximum of 19.2 Kbps data rates and will interfaces with the host computer through the PCMCIA interface adapter function.

Spread Spectrum Function: The wireless modem would be required to meet FCC part 15 in the 902-928 MHz ISM band for spread-spectrum operation. The spread-spectrum function would be addressed using a PN code generator and matched filter receiver with correlator/accumulator functions. These functions are implemented using the latest in digital signal processing technologies.

RF Function: The RF functional block uses advanced GaAs MMIC device technology as well as advanced MCMIC substrate technology. At the heart of the RF block is the Northrop PCS transceiver chip described earlier. It provides the majority of the RF functions in a 100x100 mil^2 area. These functions include the LNA, downconverter, VCO, upconverter, and .25 watt power amplifier, all operating at a nominal 3-5 volts.

GaAs MMIC provides an enabling technology for high levels of RF subsystem integration. MMICs simplify design and manufacturing for OEMs by reducing package size and weight, reducing parts inventory, and minimizing dependence on skilled technicians all while offering high levels of performance.

Recent technology advances have led to the development of MCMICs which can implement traditional lumped element components in a highly integrated, very dense package. Resistors, capacitors, inductors, and transmission lines are laid out on a ceramic substrate as-off chip circuitry for the GaAs MMIC transceiver chip. The MCMIC circuit provides for a higher level of integration when designed with Northrop MMIC

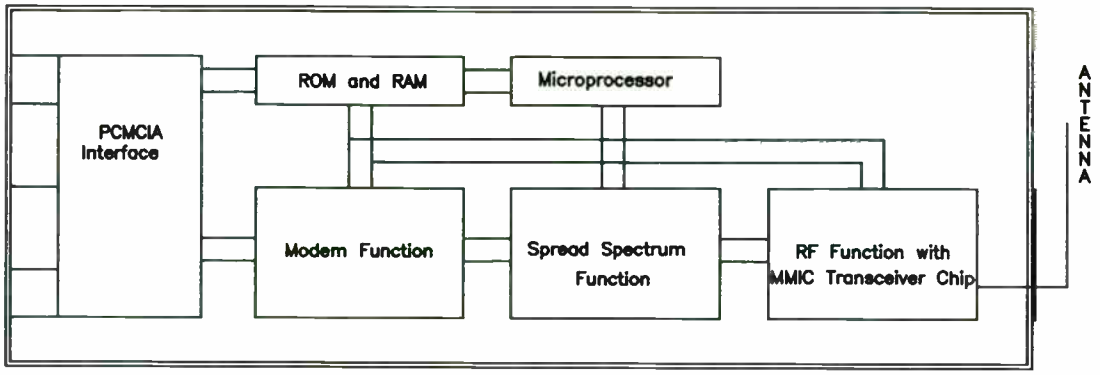


Figure 4-1. Proposed wireless modem PCMCIA card with transceiver chip.

devices.

Due to the high level of integration of the Northrop PCS transceiver MMIC, the RF layout is significantly simplified. All required components can be integrated on the MCMIC substrate as shown early in Figure 3-10. Inductors are realized with planar technology using airbridges and passivated underpasses. Capacitors can be realized with either Metal Insulator Metal (MIM) technology or interdigitated designs. This design uses all MIM capacitors. It should be noted that the MCMIC technology can be used to create a hierarchy where increased levels of integration are made possible by adding more MMIC devices to the design.

Antenna: The internal antenna provides unobstructed working space and easy portability. The design of internal antenna would take into account the PCMCIA card housing in the computer. Other external antennas would not guarantee unobstructed working space, because the location of the external antenna would change as that of PCMCIA card changes on the computer housing.

Modem Specifications: The wireless modem is designed to operate in the 902-928 MHz ISM band at a maximum data transfer rate of 19.2 Kbps. The RF design will adhere to the FCC Part 15 rules for spread spectrum operation with power levels under 1 watt. It would use a bi-phase shift-keyed modulation scheme and incorporate a PN code generator for spectrum spreading providing at least 20 dB of processing gain. The RF spectrum will occupy a 2 MHz bandwidth with a maximum of 13 separate channels. The modem design will meet PCMCIA 2.0 standards and be ExCA compatible.

900 MHz Wireless Modem Specifications

Frequency Band	902-928 MHZ
Maximum Data Transfer Rate:	19.2 Kbps
Power Level:	1 Watt max
Modulation Type:	BPSK
Access Type:	Peer-to-Peer
Module Function:	Wireless Modem
RF Bandwidth:	2 MHz
Number of Channels:	13
Processing Gain:	20 dB

5. Conclusion

A plan and development activities for a highly integrated MMIC transceiver chip for use in a PCMCIA card are presented here. The MMIC chip under development can address other PCS applications including, wireless LAN and cellular phones. The goal is to lower the cost with high volume production and inexpensive packaging. The complete package would use three technologies -- plastic, MCMIC and MMIC. The MCMIC technology contains the necessary bias circuits and the interfacing circuit between MMIC and the leads. The projected cost for the complete packaged transceiver is $30 - $40 each by the end of 1995 in volumes of 10,000 or more.

Acknowledgement

The authors are glad to acknowledge the help from Mr. Bob Anderson and Mr. Len Baran for manuscript preparations, Mr. Khoi Vu for providing measured data and Mr. Ron Langietti for useful suggestions.

LOW-POWER TRANSMITTER DESIGN USING SAW DEVICES

Earl E. Clark
RF Monolithics, Inc.
4441 Sigma Road
Dallas, Texas 75244

Surface-Acoustic-Wave-(SAW) resonator low-power transmitters have been in widespread use since the early 1980s. They are found in a wide variety of applications such as automotive keyless entry, door and gate openers, wireless alarm sensors, medical alert pendants, bar code readers, and many other applications in the wireless remote control, security, and data transmission area. There are many constraints that impact the transmitter configuration: range, data rate, battery life, size, and emissions requirements. The inter-relationship of these constraints must be understood so that an optimal transmitter is designed. This paper will focus on the different types of SAW-resonator low-power transmitters and their performance characteristics.

There are many benefits to using SAW resonators in low-power transmitters. They provide outstanding performance when compared to traditional LC stabilized transmitters, frequency synthesizers, or multiplied bulk crystal transmitters. The SAW resonator-based oscillator, used in low-power transmitters, is a very stable, fundamental mode frequency source at UHF. Properly designed SAW oscillators are relatively insensitive to changing load impedance and have good temperature stability. They are very rugged and, being hermetically sealed in TO-39 package, have excellent aging characteristics. Being a fundamental mode devices means the circuit complexity is greatly reduced. This has a direct impact on overall transmitter size. SAW devices simplify product design and manufacturing by removing costly alignment steps. These properties give SAW resonator-based oscillators a very low cost/performance ratio.

In order to gain the maximum benefit from using a SAW device in the transmitter it is important to take advantage of the SAW device's characteristics when designing a new system. Due to their fundamental mode of operation and high degree of temperature stability, SAW devices can be used to stabilize the frequency of both the receiver and the transmitter. Therefore, receiver bandwidths can be reduced, increasing sensitivity and decreasing susceptibility to interfering signals. Due to the explosion of wireless applications, the finite amount of spectrum allocated to low-power applications is becoming more and more crowded. The use of SAW devices, whether resonators for frequency stability, or filters for rejection of out-of-band signals, is a very cost effective way to decrease the required system bandwidth.

Using SAW devices in a system reduces the circuit complexity. This means that systems can be made smaller without sacrificing the performance advantages of more complex systems. Additionally, a less complicated circuit results in a lower power requirement, conserving precious battery resources. SAW devices make it possible to have a simple, low-power system, with no production alignment, and performance characteristics that rival much more complex systems.

SAW resonators are fabricated by depositing a thin film of metal, typically aluminum, onto a highly polished quartz substrate. The frequency, Q, and insertion loss are a function of the geometric pattern that is etched into the metal. Figure 1 shows a drawing of the different parts of a resonator.

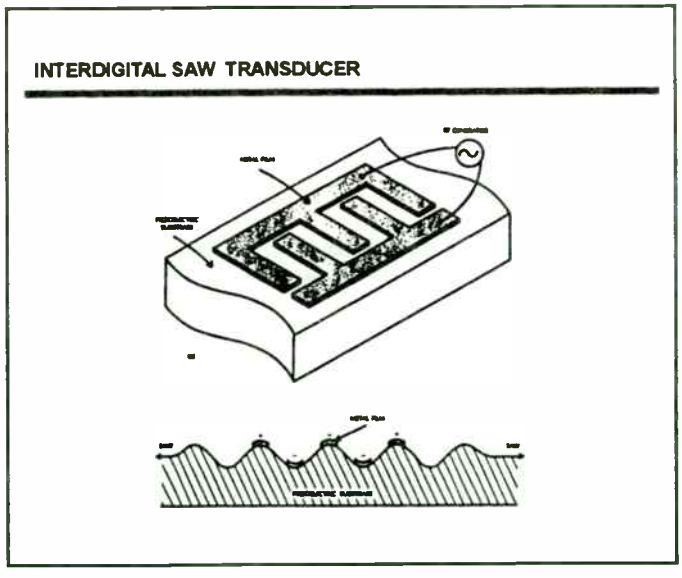


Figure 1: SAW Resonators

SAWs can be fabricated on many types of piezoelectric substrates. There are Lithium Niobate and Lithium Tantalate used for wide band SAW filters, and ST cut Quartz used for narrow band SAW filters and resonators. Of these, quartz is the most temperature stable. Figure 2 shows the relative temperature stability of the three materials.

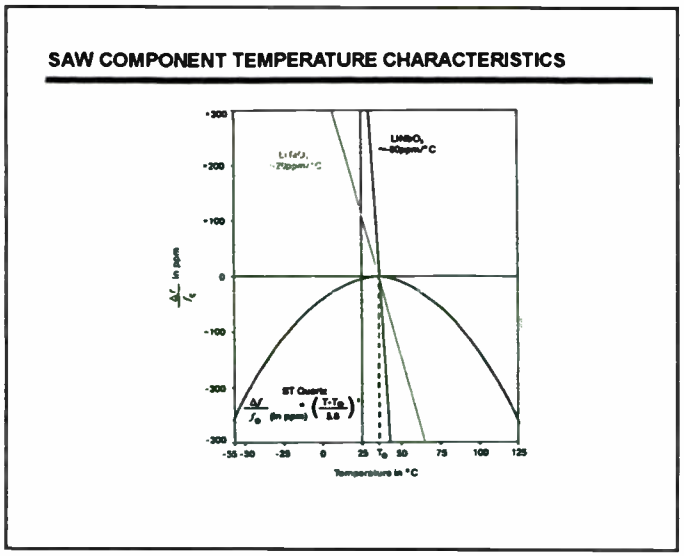


Figure 2: Temperature Characteristics

The frequency of the SAW resonator and oscillator exhibits the parabolic dependence described by the equation in the figure. For a typical transmitter, designed to operate from -40C to +85C, the change in center frequency would be approximately 125 PPM. This is at the minimum when the SAW resonator has its turnover temperature centered in the operating temperature range. The turnover temperature is defined where the frequency is at the maximum value and is set by the design of the SAW device.

SAW resonators can be modeled by an equivalent circuit using lumped elements. Figures 3 and 4 show the equivalent circuit for two commonly used SAW resonators in low-power transmitters. Figure 3 is the equivalent circuit for a two-port resonator. This model is valid near the center frequency of the SAW resonator. The phase shift through this device can be set to either 0 or 180 degrees. The equivalent circuit for the more popular one-port resonator is shown in Figure 4.

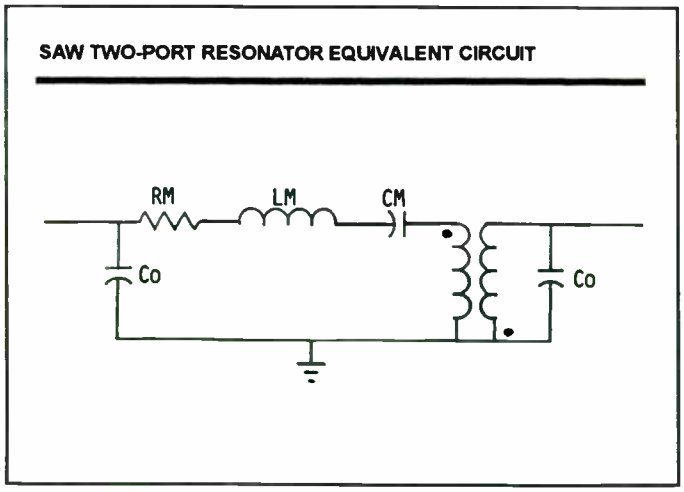


Figure 3: Two-port Equivalent Circuit

These one-port resonators are used as the frequency determining element in a Colpitts oscillator. This topology is

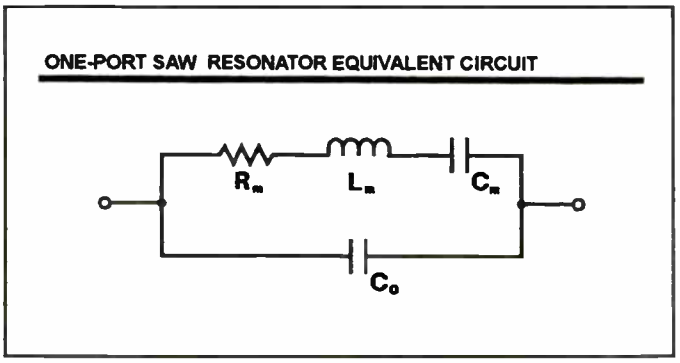


Figure 4: One-port Equivalent Circuit

the most popular found in low-power transmitters. Figure 5 shows the reactance versus frequency of a single-port SAW resonator. This plot can be used to determine the frequency of oscillation when the circuit is tuned properly. The frequency of oscillation will be at the low reactance point near the real axis. The exact frequency will depend on the circuit tuning.

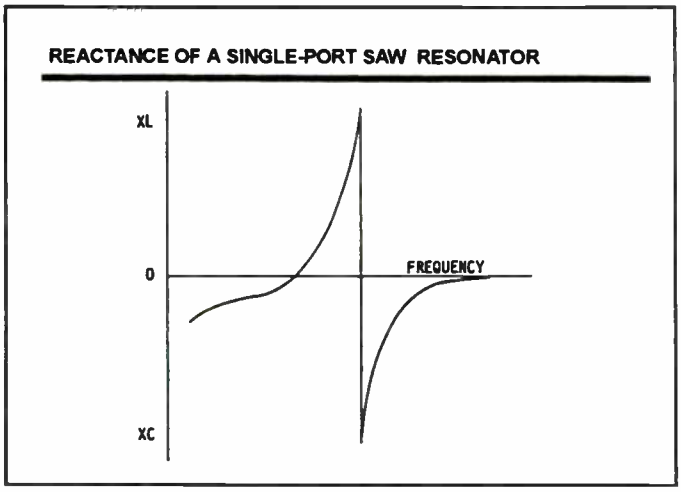


Figure 5: Reactance versus Frequency

Figure 6 shows the schematic diagram of the first generation SAW transmitter. The circuit uses the Colpitts oscil-

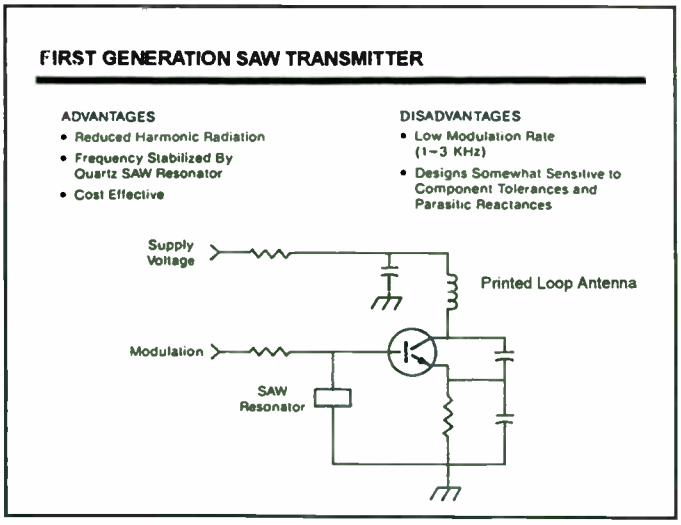


Figure 6: First Generation SAW Transmitter

LOW POWER SYSTEM REGULATIONS

COUNTRY	REGULATION	FREQUENCY LIMITS (MHz)	FUNDAMENTAL POWER	2nd HARMONIC	3rd HARMONIC	GREATER THAN 3rd
UNITED KINGDOM	DTI MPT1340	417.9 - 418.1	250 uW	4 nW	1 uW	1 uW
GERMANY	FTZ 17 TR 2100	433.05 - 434.79	25 mW	1 nW	30 nW	30 nW
FRANCE	CNET PAA1542	224.6 - 224.8	5 mW	250 nW	4 nW	250 nW
NETHERLANDS	PTT	433.052 -434.797	50 uW	4 nW	1 uW	1 uW
JAPAN	MPT	303.675-303.975	500 uV/m	35 uV/m	35 uV/m	35 uV/m
U.S.	FCC PART 15	260 - 470	3750-12500 uV/m	375-1250 uV/m	375-1250 uV/m	375-1250 uV/m

Figure 7: Regulations

lator topology. The printed loop antenna is used as the tuning inductor with the feedback accomplished with the collector to emitter capacitor. The feedback should be set so that the gain of the circuit is not so great as to cause oscillation at frequencies other than the SAW frequency, or so small that the oscillator does not modulate properly.

There are several advantages to this circuit such as no production tuning, minimum parts count, and the ability to be designed so that the harmonic emissions meet the FCC Part 15 specification. This results in a very cost effective, high-performance transmitter.

This generation of transmitter has proven to be both popular and reliable for applications that fall under FCC Part 15. There are some limitations, however, to this design. Due to the high Q of the resonator, the data rate is limited to a maximum of about 3 kHz. The transmitter frequency can also be sensitive to component tolerances and parasitic reactance. An additional problem encountered is trying to use this circuit topology for a transmitter that will be subject to the European emissions regulations. These regulations are much more stringent in the allowable level of the harmonic than the FCC Part 15. The chart shown in Figure 7 highlights the important differences in the various regulations.

The series-fed antenna, typically a trace on the circuit board, does not provide enough frequency selectivity for the transmitter to meet the more stringent regulations. There are some design tools that can help such as using a lower frequency transistor, which further reduces the maximum data rate, or using a trap at the second harmonic which typically calls for an alignment step in manufacturing. Both of these options set further constraints on the transmitter design.

In order to comply with the European regulations, specifically DTI MPT1340, RFM developed the MB1005. (This circuit is shown schematically in Figure 8.) Like its predecessor, it also uses a Colpitts oscillator; but in order to reduce the harmonic emissions to the required level, two modifications have to be made. First, the fundamental output power is reduced so that the required rejection of the second harmonic is not as great. Second, the addition of two adjustments is necessary. One adjustment is needed to set the oscillator on frequency and the other is needed to tune the antenna. This design makes use of a tapped antenna, which has a higher impedance allowing the Q of the tuned antenna circuit to be made higher. This reduction of fundamental power and use of a higher impedance antenna is typical of most of the transmitter designs used in the European market.

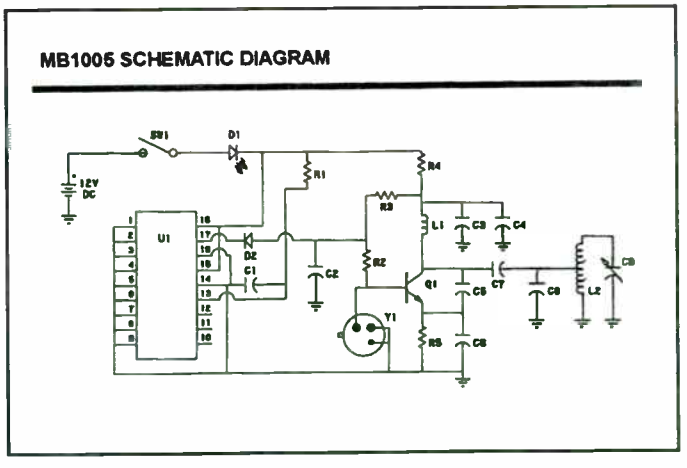


Figure 8: MB1005

However, even this design was not capable of meeting the German FTZ 17 TR2100 emissions requirements. A further reduction in output power would be required for this design to pass the more stringent second harmonic specification. This would obviously have a great impact on the system range.

The need for greater power, reduced harmonic levels, and no adjustments gave rise to a new design approach for low-

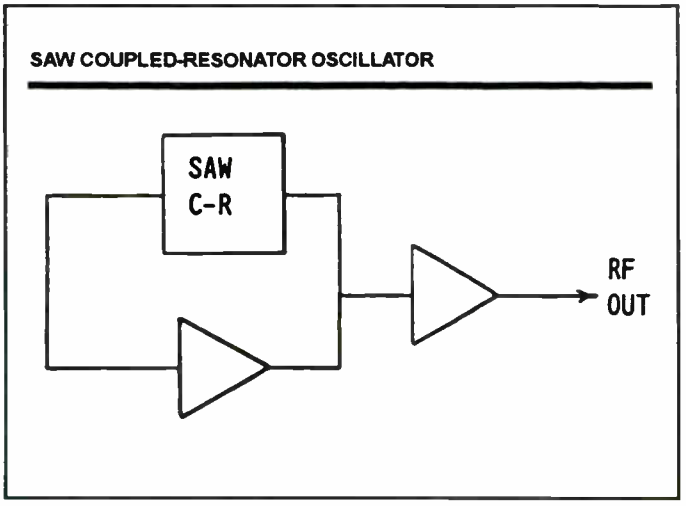


Figure 9: Pierce Block

power transmitters. This design uses a Pierce oscillator configuration which is shown in the block diagram of Figure 9.

In order to implement this circuit, RFM developed a SAW coupled resonator. The coupled resonator is a two-pole SAW resonator with a frequency response similar to that shown in Figure 10. This is contrasted with Figure 11 which shows the response of a standard resonator at the same frequency. Perhaps the most important difference between the two devices is the amount of phase shift across the 3 dB bandwidth. The standard resonator having a single-pole has only 90 degrees of phase shift while the two-pole coupled resonator has 180 degrees.

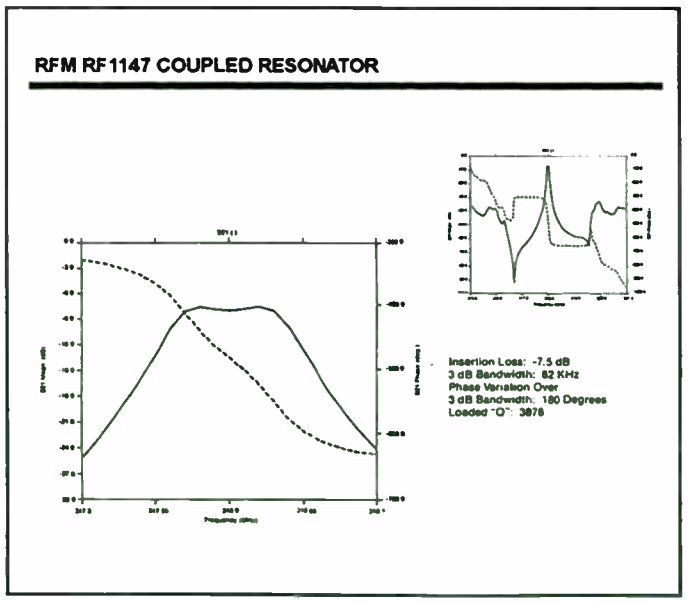


Figure 10: Coupled-resonator

Referring to the block diagram in Figure 9, the requirement for oscillation is that the gain around the loop be at least unity with a phase shift of 0 degrees, 360 degrees, or an integer multiple of 360.

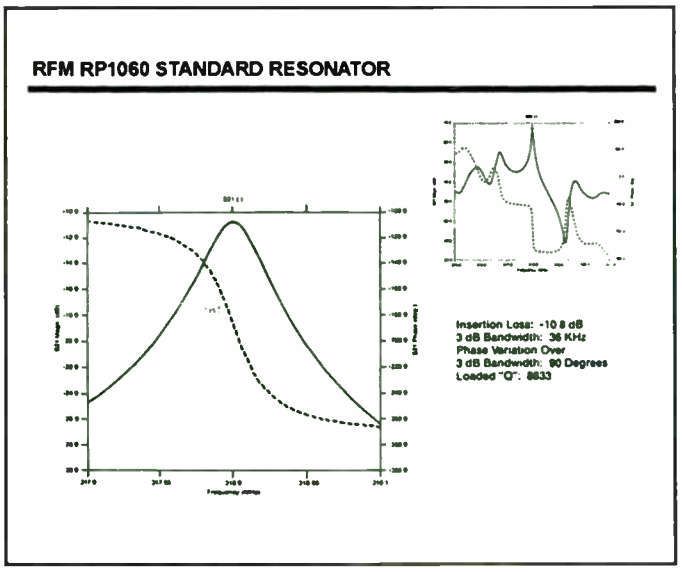


Figure 11: Two-port Resonator

The importance of the extra 90 degrees of phase shift across the SAW coupled-resonator bandwidth can be illustrated with the following example:

The phase shift through a normal transistor amplifier at UHF frequencies is approximately 240 degrees. In order for this circuit to meet the phase requirement, the phase shift provided by the feedback element must be either +120 or -240 degrees. The normal single-pole resonator has either 0 or 180 degrees of phase shift at the center frequency with a +/-45 degree phase shift across the 3 dB bandwidth. A quick calculation shows that the phase criteria would not be met within the 3 db bandwidth of the resonator.

The coupled resonator, on the other hand, has a similar absolute phase at the center frequency, but has a +/-90 degree phase shift across the 3 dB bandwidth. Using a coupled resonator and a properly designed amplifier, it is possible to guarantee the oscillator frequency to be within the 3 dB bandwidth of the SAW. The benefit of this topology is that it requires no tuning and is relatively insensitive to stray reactance. In contrast, if a single-port resonator were used it would be necessary to introduce an added phase-shift circuit to the oscillator loop in order to obtain the required phase shift.

Based on this approach, RFM then designed its first integrated circuit transmitter called the Microtransmitter (MX). A block diagram of the MX is shown in Figure 12. This circuit is designed for FCC Part 15 applications and is capable of full compliance with all current and proposed FCC Part 15 regulations.

A buffer amplifier is used so that the oscillator frequency is insensitive to changes in load parameters. The buffer amplifier also allows for a much higher modulation rate than would be possible if the oscillator were to be directly

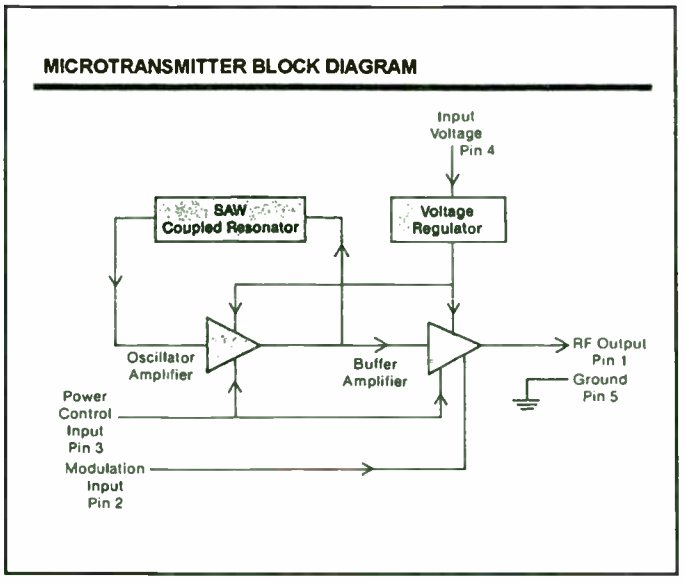


Figure 12: MX Block Diagram

DESIGN GOAL PERFORMANCE AND MEASURED PERFORMANCE OF MICROTRANSMITTER AT 318 MHz

Specification	Design Goal	Typical Measured Performance
1. Output Frequency	318.000 MHz +/- 250 KHz	318.000 MHz +/- 100 KHz
2. RF Output Power (50Ω Load)	≥ +7 dBm	+12 dBm
3. Power Output Variation over a Supply Voltage Range 6.0 to 10.0 V_{DC}	+/-1 dB	+/-1 dB
4. Modulation Depth (On/Off Ratio)	35 dB Min	>50 dB
5. Modulation Rate Capability	<50 KHz	<50 KHz
6. Modulation Rise or Fall Time	<1 µsec	<1 µsec
7. Radiated Harmonics when used with RFM Specified Antenna	-20 dBc	-20 dBc
8. Output Power Control Adjustment Range	15 dB	10 dB
9. Power Supply Current Drain at 10 V_{DC} and Maximum Power	<25 mA	<20 mA
10. Power Supply Current Drain at 10 V_{DC} and and Power Control Resistor Grounded	<1 µA	<10 µA

Figure 13: MX Performance

modulated. The design goals and actual performance are shown in Figure 13.

A transmitter based on the MX is capable of putting out the maximum amount of power allowed and still meeting the emissions specification under FCC Part 15. It also is capable of a high data rate, up to 50 kHz, and requires no production tuning. The disadvantages of this product are its power consumption, which limits its battery powered applications, and its relatively high harmonic levels, which make it difficult to use in European applications.

The MB1003 demonstration transmitter was developed to show that a European design could be accomplished with the Microtransmitter. Although it requires an adjustment to center the antenna circuit, it reduces the number of adjustments from two, on the MB1005, to one. The schematic representation of this circuit is shown in Figure 14.

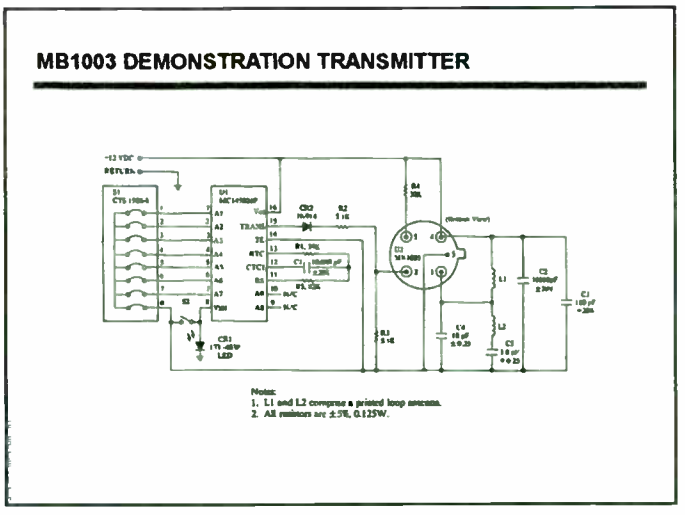


Figure 14: MB1003 Schematic

In order to more effectively address the unique emissions requirements of the European market as well as US markets that require a high degree of miniaturization, RFM has developed the third generation of transmitter components: the HX series. This series is a fully functional RF building block that engineers, who may have little RF background, can incorporate into a transmitter design with little effort.

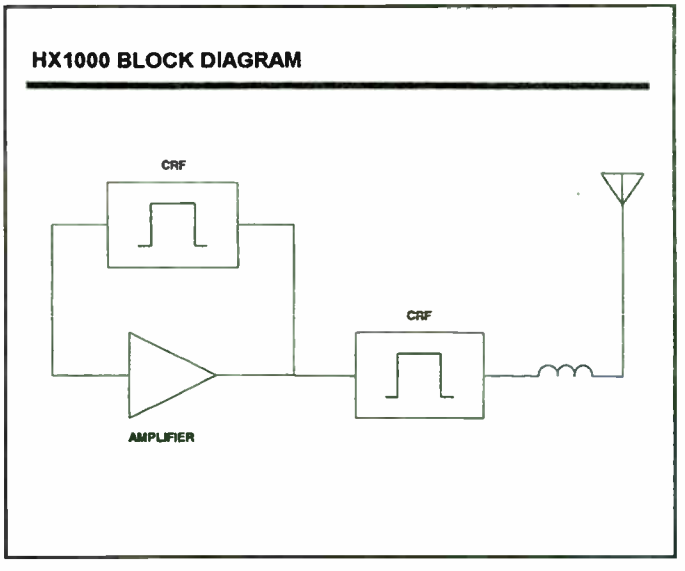


Figure 15: HX Block Diagram

The HX is a hybrid transmitter that is packaged in a hermetically sealed surface mount package. The block diagram of the HX is shown in Figure 15. The oscillator topology is a Pierce configuration using a SAW coupled resonator as the feedback element. Like the Microtransmitter the HX transmitter is also very insensitive to changing load conditions and parasitic reactance.

Advances in SAW technology at RFM have led to coupled resonator designs that have 3-5 dB of insertion loss, untuned, in a 50 Ohm test fixture, at frequencies as high as 930 MHz. This new SAW technology has greatly simplified the oscillator circuit. The result being a very

cost effective transmitter building block. This advanced low-loss coupled resonator has also been used as the output filter. With this new SAW coupled resonator, the fundamental output power can be high and still provide the necessary harmonic rejection. It was necessary to have a low-loss filter so that the transmitter power could be made as high as possible with a minimum amount of current, thus enhancing battery life.

MB1007 AM KEYHEAD TRANSMITTER (Performance Parameters)

Characteristics	Minimum	Typical	Maximum	Units
Operating Frequency	433.72	433.92	434.12	MHz
RF Output Power	25	50		µW
Spurious Emissions		FTZ 17TR2100 Compliant		
Power Supply	2.7	3	3.3	Volts
Operating Current		7	10	mA Peak
Operating Temperature	-30		+85	C
Data Rate		1		KHz
Oscillator Turn On Time			100	µs
Oscillator Turn Off Time			100	µs

Figure 16: HX Performance

Figure 16 shows the specifications of the HX1000: the hybrid transmitter designed for the German market. This module, when properly used, will meet the German FTZ 17 TR2100 emissions specification. This is evidenced by the MB1007 demonstration unit whose data is shown in Figure 17.

MB1007 TRANSMITTER PERFORMANCE vs REGULATION (Handheld Measurements)

Harmonic	Freq (MHz)	Power Measured (dBm)	Power Measured (Watts)	FTZ Spec (Watts)	DTI Spec (Watts)
Fund.	433.92	-12	63E-6	25E-3	250E-6
2nd	867.84	-60	1E-9	1E-9	4E-9
3rd	1301.76	-65	0.3E-9	30E-9	1E-6
4th	1735.68	-62	0.6E-9	30E-9	1E-6
5th	2169.70	-46	25E-9	30E-9	1E-6
6th	2603.52	-49	13E-9	30E-9	1E-6
7th	3037.44	-47	20E-9	30E-9	1E-6
8th	3471.36	-48	16E-9	30E-9	1E-6

Figure 17: MB1007 Performance

The HX series of hybrid transmitters are designed to operate on a 3 volt Lithium battery. This greatly reduces the size of the transmitter. The MB1007 demonstration transmitter is configured on the head of an ignition key and uses the key stem as the antenna. In addition to the HX1000 at 433.92 MHz, an HX has been designed for all the other major low-power frequencies.

Conclusion

SAW resonator transmitters have been around for many years. The differing governmental regulations have given rise to the need for the various transmitter topologies. In the US, the harmonic emission requirement is only 20 dBc. This leads to a simple low-cost design. The major changes in these transmitters has been the replacement of the LC frequency determining elements with the SAW resonator. The surface mount revolution is currently causing changes in transmitter design. SAW resonators in surface mount packages are now available.

The European requlations, with more stringent harmonic requirements, create the need for a different transmitter design. These harmonic requirements generally cause the transmitter power to be low or have an elaborate filter. New advances in SAW technology at RFM have led to RF modules that allow for both a high fundamental power level, and sufficient harmonic rejection to meet the appropriate regulation.

Acknowledgment

The author wishes to thank Christina McFarland for her help in putting this paper together.

FILTER COMPARATOR NETWORK FOR BEAM POSITION MONITORING

Michael Ferrand and Mark McWhorter
Vernitron Corporation
Lorch Electronics Division
2801 72nd St. N. St. Petersburg, FL 33710

Lorch Electronics has completed development and is currently manufacturing a Filter / Comparator network has been developed for Argonne National Laboratory for use in their Advanced Photon Source. This passive device operates in close proximity to the accelerator ring and provides boresight accuracy measurements to aid in the alignment of the beam.

The Filter / Comparator is the front end of a Beam Position Monitor, and is installed in close proximity to the particle accelerator ring. The device is connected by short phase matched cables to four capacitive "buttons" inserted on the perimeter of the ring which produce a very sharp voltage spike each time a particle "bunch" passes by. See Figure 1. The signal is an impulse 30 to 60 picoseconds wide, with an amplitude of 0.2 to 200 Volts. These signals must be stretched to about 100 nanoseconds in order to allow processing.

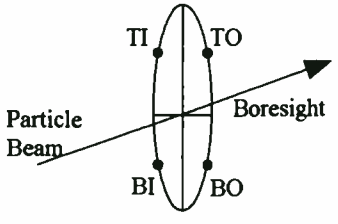


Figure 1 - Boresight

The purpose of the Filter / Comparator is to stretch and compare the signals from the four buttons in the time domain, such that the Sum and the Differences of the X and Y axis of those signals are provided with a high degree of accuracy.

The stretching is performed by matched sets of Gaussian response filters. The arithmatic summing / differencing is performed by the Hybrid Matrix. Additional Test / Trigger circuitry for timing is realized using Directional Couplers and a 4-way Power Splitter.
The beam position within the ring is related by the amplitude of the device outputs. Time domain distortion of the amplitude in terms of overshoot and ringing may create a measurement error.

The Filter / Comparator is comprised of the following matched components:

4-Input Attenuators, 6 dB, Broadband
4-Directional Couplers, 10 dB, Broadband
1-Power Divider, 4-Way, Broadband
3-Bandpass Filters, 352 MHz.
4-Lowpass Filters, 400 MHz.
4-180° Hybrids, Broadband
4-Amplitude / Phase Trim Circuits (as required)

Figure 2 provides a block diagram of the Filter / Comparator. The four inputs are labeled TI (Top In), TO (Top Out), BI (Bottom In) and BO (Bottom Out).

The individual components that comprise the Filter / Comparator are described below.

INPUT ATTENUATORS

The input attenuators reduce the signal by 6 dB at the input. Their primary function is to avoid the development of large standing waves caused by the reflections between the buttons and the

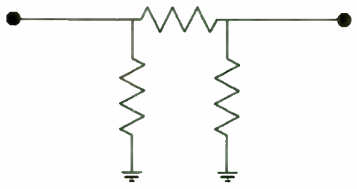


Figure 3 - 'Pi' Pad

TEST / TRIGGER CIRCUIT - 10 dB DIRECTIONAL COUPLERS / POWER SPLITTER

The couplers combine the reflected signals from the filters, and create a trigger for timing purposes. They also allow the injection of four test signals into the system. The coupled port on BO will have a 6 dB pad added to create a predetermined offset, negative for Dx and positive for Dy. The path from the 4-way power divider input to each coupler output is matched in amplitude and phase. The couplers are comprised of lumped elements to allow for adjustment in matching amplitude and phase.

BANDPASS FILTERS

The Filter / Comparator contains three bandpass filters whose purpose is to provide a Gaussian shaped 100 nanosecond wide modulated pulse from the input impulse passing through the lowpass filters and hybrid matrix. The filters must be matched in amplitude, phase and delay. So that the filters were matched well beyond their 3 dB points a 12 dB "Transitional Gaussian" filter was used. The individual filter specifications were as follows:

Center Frequency: 351.93 MHz
3 dB Bandwidth: 10 MHz
Number of Poles: 3
Sidelobe Rejection: 60 dB. min.
Amplitude Matching: ±0.05 dB
Phase Matching: ±0.5°
Insertion Loss: 4 dB max.

With the exception of the matching requirements, the filter fell into the realm of a "catalog standard". The filter was realized as a mutually-coupled circuit which provides excellent insertion loss and ease of alignment. The filter schematic is shown below in Figure 4.

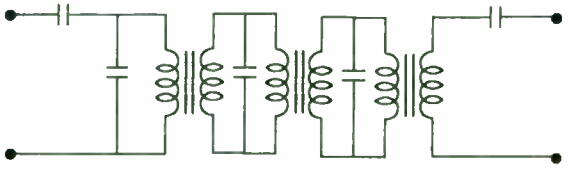


Figure 4 - Bandpass Filter

180 DEGREE HYBRID BUTLER MATRIX

The Filter Comparator uses four 180° Hybrids to linearly process the outputs of the filters in the time domain. The hybrid matrix produces three output signals:

SUM=TI+TO+BI+BO
DY=TI+TO-BI-BO
DX=TO+BO-TI-BI

The 180 degree hybrids are constructed using twisted pair ferrite miniature transformers and have a usable bandwidth from 10 to 500 MHz. In order to meet the key specification of cancellation ratio (-45 dB @ Dx and Dy outputs with inputs at boresight), we tuned the hybrids to less than one degree phase error and 0.15 dB amplitude error between all four ports at the critical frequency of 351.93 MHz. Sets of four phase / amplitude matched hybrids are required to create the matrix. The fourth output of the matrix is terminated internally into 50 Ohms. We manufactured the hybrids were manufactured on a PC mountable header. A schematic of the 180° hybrids is shown in Figure 5.

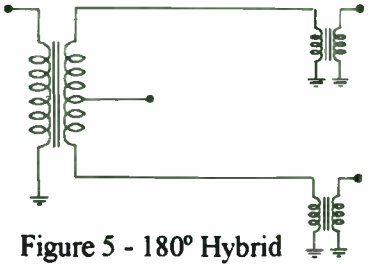


Figure 5 - 180° Hybrid

LOWPASS FILTER

The lowpass filters are required to reject the high frequency components of the input impulse to the Trigger output, and allow the desired frequency of 351.93 MHz to pass for processing into the hybrid matrix. They are realized as a 3 pole lumped element Chebyshev design. The series inductor is tuned during the final alignment process to optimize the cancellation ratio.

AMPLITUDE / PHASE TRIMMERS

In order to achieve the 45 dB cancellation ratio, we must compensate for manufacturing and component tolerences. An adjustment is provided by the lowpass filter's series inductor. Additional capacitive tuning between hybrids of the Butler matrix is also required. In order to phase and amplitude match all three outputs, it is necessary to use "select at test" shunt chip resistors and capacitors at each filter output.

MECHANICAL CONFIGURATION

The components are mounted on a PC board inserted into a machined aluminium housing. Because of the necessity of maintaining equal phase lengths / delays between channels, all the RF traces on the board are the same length. The PC board is manufactured from FR-4.

The system application is in close proximity to the accelerator ring. The Filter / Comparator is designed to withstand a high gamma radiation environment of up to 50 Krads/year, which precludes the use of radiation sensitive materials within the unit. Teflon® could not be used, therefore the connector dielectric is made of Rexolite. The epoxy to secure the components is also certified to withstand this type of environment.

The finished unit is 6 x 6 x 2 inches. SMA-female connectors are used for the TI, TO, BI, BO inputs and TNC female connectors are used for the Sum, Dx, Dy and Test/Trigger

THEORETICAL DATA

The customer required exhaustive computerized analysis to show the effects of component tolerances. We presented design data and analysis to the customer in December of 1992. The report contained well over one hundred pages of test data and drawings and showed the effects of component tolerance on the cancellation ratio. Data on individual components including the coupler, pads, filters, hybrids and the power splitter were also provided.

The customer required data to show performance in both the time and frequency domains. For analysis in the time domain we used iSpice by Intusoft. For the frequency domain, we used Touchstone® by EEsof was used.

In addition to analyzing the individual components, we performed a complete systems analysis in the time and frequency domain.

ACTUAL DATA

The key performance criteria of the device is cancellation ratio. Return loss and insertion loss / phases are also measured on each unit.

Cancellation ratio is defined at boresight in the ring, where all four input signals have identical amplitude and time of arrival at the device inputs. Under this condition, each device must be painstakingly tuned to achieve the desired -45 dB ratio of Sum to Dx (or Dy) output.

Tests were carried out using both methods until it could be proven that correlation existed between Time and Frequency Domain measurement techniques. Production testing was then performed using the HP-8753C network analyzer with the Time Domain Option. The units are tuned in the frequency domain while simultaneously viewing both D-output ports. Final data is taken using the Time Domain option.

Figure 6 shows the smooth Gaussian shape of the Sum output, which is the sum of all 4 inputs.

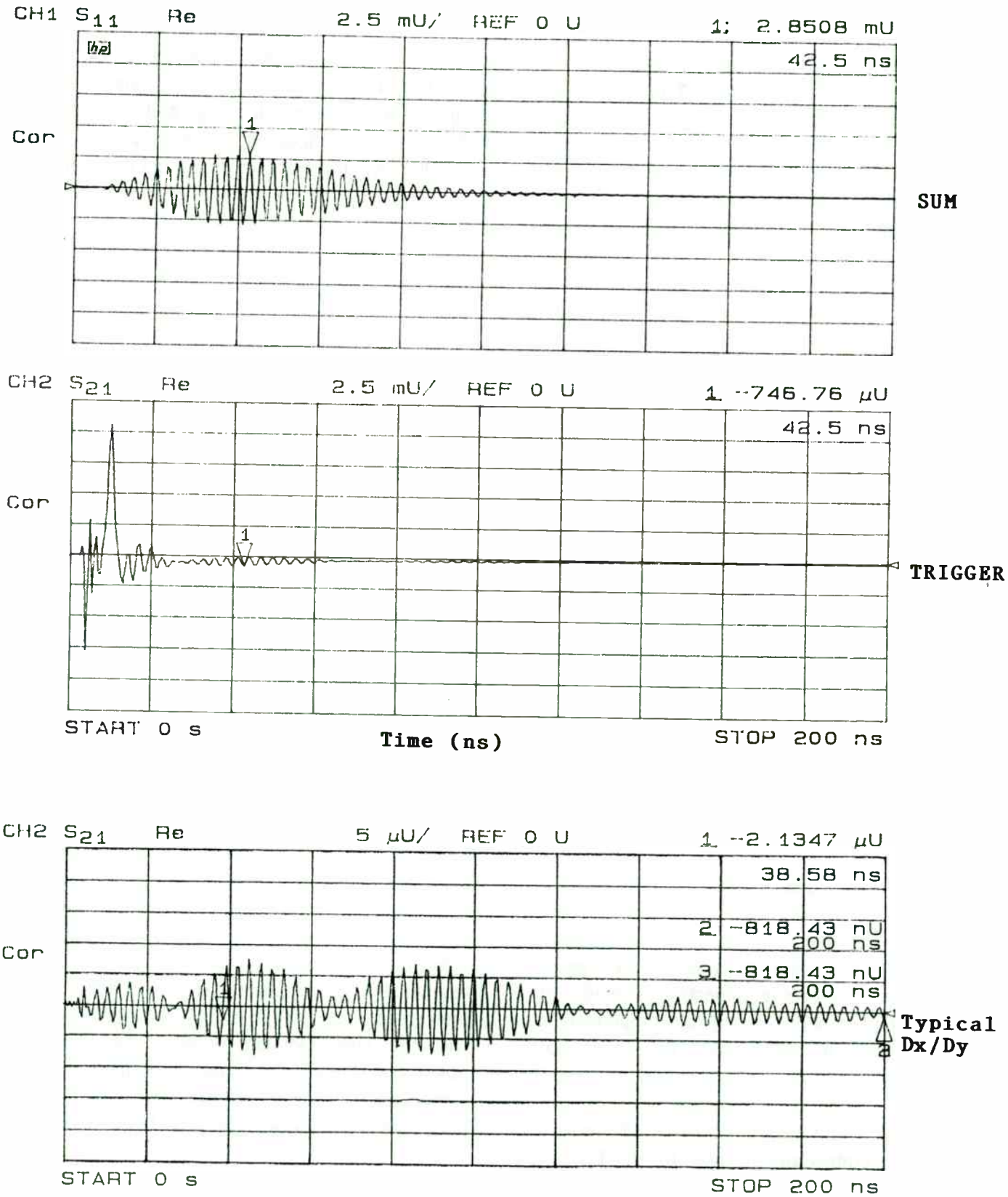


FIGURE 6

It also shows the typical Dx (or Dy) output at -45 dB or more and the narrow pulse signal, used for timing at the next stage of the Beam Position Monitor.

Measured into a fifty ohm system, the Filter Comparator has typically 13 dB of insertion loss, and phase balance between outputs of ±3 degrees with various input conditions applied. Return losses are typically 16 dB.

Lorch is currently producing five hundred units in support of the Argonne Advanced Photon Source.

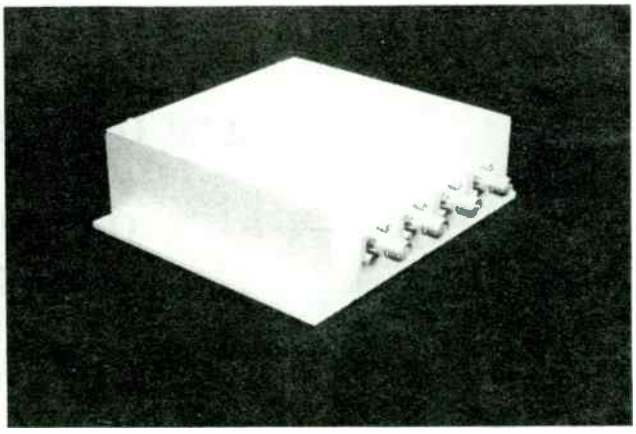


Figure 7. Filter/Comparator

This work was carried out under a U.S. Department of Energy contract through Argonne National Laboratories. The authors would like to acknowledge the design effort and help of Manny Kahana of Argonne National Laboratories and Tho Van Nyugen for his bench work and support.

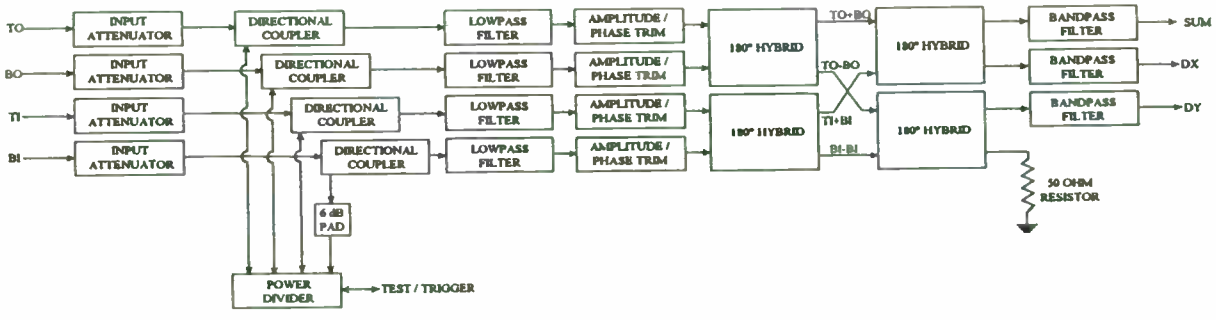


Figure 2

BLOCK DIAGRAM, FILTER COMPARATOR UNIT

Digital Temperature Compensation of Oscillators Using a Mixed Mode ASIC

Steven Fry
muRata ERiE North America
1900 West College Avenue
State College, Pa. 16801

Abstract

This paper will describe the design, construction, calibration and operation of digitally compensated crystal oscillators, making use of a custom Application Specific Integrated Circuit developed specifically for the purpose. This IC is fabricated in a "Triple Technology" process. This allows the integration of all of the digital functions, supporting analog blocks and an Electrically Erasable PROM in a single +5 V device. A true "single chip" DCXO is therefore made possible.

Also described will be a novel compensation scheme which allows autonomous calibration, requiring no interface to an external test computer in order to acquire the unique set of compensation data for each oscillator.

Introduction

Digital frequency vs. temperature compensation has been a reality since the early days of the integrated circuit. Even then, these schemes achieved quite respectable performance with some capable of achieving compensation less than the magnitude of typical quartz crystal thermal hysteresis. (1)

These original devices tended to be bulky and complicated to assemble and were therefore large and expensive. As the available integrated circuits began to improve and with the aid of hybrid construction techniques, the oscillator sizes began to shrink, but the costs remained relatively high. (2)

With the advent of Large Scale mixed mode integration, it is now feasible to design a true "single chip" DCXO allowing dramatic size and cost reductions.

Digital Compensation Basics

The block diagram of a fundamental digital compensation system is shown in Figure 1. A temperature sensor which is in close thermal contact with the crystal senses the instantaneous crystal temperature. This analog signal is then properly scaled to match the end points of the operating temperature range to the zero and full scale limits of the A/D converter. Almost any type of A/D may be used since conversion speed is not an issue. In some cases, a temperature to frequency converter using a crystal may be used.

The output of the A/D then becomes the address which is applied to a non-volatile ROM which contains the required compensation data at each temperature increment. This memory is typically a low-powered CMOS device.

The contents of the addressed memory location are then applied to the D/A converter which produces the compensating voltage which is fed back to the voltage control port on the crystal oscillator to compensate the frequency. This type of "brute force" look-up table technique has been used successfully by a number of manufacturers, but the assembly is still relatively complicated and a coarse compensation network is sometimes necessary to reduce the frequency excursion of the crystal to a lower level.

The ultimate achievable accuracy of the system is determined by several factors: (3) (4)

1.) The total temperature range to be covered.
(T=Tmax - Tmin)

2.) The maximum peak-to-peak frequency excursion of the crystal.
(F=Fmax - Fmin)(PPM)

3.) The maximum slope of the crystal's frequency vs. temperature characteristic.
(S=PPM/Deg. C)

4.) The number of bits of resolution of the temperature measurement as determined by the A/D.
(Nt=# of A/D bits)

5.) The smallest increment of frequency adjustment as determined by the D/A.
(Nv=# of D/A bits)

The smallest delta F achievable can then be calculated by:

$$dF= (S/2)(T/2^{Nt}) + (F/2)(1/2^{Nv})$$

One additional factor which must be taken into account is that this formula assumes that 100% of the A/D and D/A range will be used. It is more practical to expect that an average of 70% will be usable without spending a large amount of test time on the initial set-up, therefore:

$$dF= (S/2)(T/(0.7)(2^{Nt})) + (F/2)(1/(0.7)(2^{Nv}))$$

The typical calibration procedure for this type of DCXO would require that the PROM be replaced with an interface cable which would be connected to a test computer. As the Unit Under Test is then put through a temp. run, the computer reads the A/D and exercises the D/A to determine the required data. A curve-fit is then performed and the entire data table is burned into the PROM. (5) This procedure is fairly straight forward, but designing a system to compensate multiple units can be quite complicated. It would therefore be desirable if some on board intelligence could be added to the oscillator to simplify the data acquisition for calibration. Several companies have taken this approach by adding a micro-computer to the oscillator, but the procedure still requires interactive communication with the host computer and assorted instruments. It was therefore a goal of the project undertaken by muRata to produce an oscillator which was "self compensating".

ASIC Configuration

With the goals of producing a single-chip DCXO that was cost effective and required a minimum amount of testing time and test hardware, the necessary block diagram was designed. After some refinements to match the configuration to the chip vendor's processes and cell library, the system was breadboarded using similar packaged kit parts. The circuit was controlled with an emulator which also served as the software development tool. After finalizing the hardware configuration and the system software for the internal ROM, the chips were produced by Sierra Semiconductor.

Digital Section

All of the chip operations are controlled by a COP880 series 8-Bit microcomputer cell. The memory available consists of: 128 Bytes of RAM

for temporary program storage; 4 KBytes of mask programmed ROM which contains the firmware to execute the self-calibration and normal operational modes and 256 Bytes of EEPROM into which the uCPU loads the compensation data as it is acquired. A charge pump is included so that write operations to this non-volatile memory are completely transparent. Analog cells are used to implement the power-on-reset and low voltage detect functions which ensure reliable operation of the uCPU under any power supply conditions.

The clock for the digital section is derived from the external crystal which is prescaled by an on-chip divider. This divider is adjustable by the uCPU from /5 to /127. The desirable clock frequency to the uCPU is around 620 KHz so that a wide range of external crystal frequencies may be used directly.

A serial address/data bus allows communication with an external computer or test fixture to load the application specific program constants, to run diagnostics or to examine the data tables after calibration. Using this bus, the uCPU can also execute from an external ROM so that the chip operations maybe customized for certain lower volume applications without changing the on-chip mask ROM and fabricating new wafers.

Also included in the digital section is a power-down timer. The uCPU may load this register with a specific delay time and then put itself into a "Halt" mode. Program execution is then suspended until the timer period has elapsed and the uCPU is re-started. This halt period is dynamically adjusted by the uCPU as it monitors the rate at which the temperature is changing. When the temperature is stable, the compensation data output only needs to be updated periodically and the "sleep" mode may extend to greater than 2 minutes to conserve power and reduce the noise generated. An external signal may be asserted to completely inhibit the uCPU from running if so desired.

Several general purpose digital I/O pins are available for monitoring program status, accurate signal timing, or any digital I/O function as defined by the program. A programmable pulse counter is also available which may be used for synthesis of accurate low frequency signals.

Analog Section

The measurement of the crystal temperature is implemented with 12-Bits of accuracy by a dual slope integrating A/D converter. This type of A/D was chosen for the most efficient use of die area, low power consumption and capability of achieving the needed accuracy. This configuration does require two external capacitors and a resistor but good performance has been achieved with small, low cost components. The integration cycles are controlled and switched by the uCPU which also counts clock cycles to perform the conversion. Although the conversion time is typically 130 mSec, speed is of little concern here since the parameter being measured has a relatively slow rate of change.

The temperature of the ASIC die is indicated by an on-chip temp. sensor. If closer thermal contact with the crystal is necessary, an analog switch is set so that the voltage from an external temp. sensor is applied to the A/D amplifier. This amplifier must be a low noise device since the equivalent bit size referred to the input may be less than 10 uV.

The temp sensors typically produce a voltage of 300 mV at +25 C and change at +1 mV/Deg. C. The A/D's zero and full scale points are then scaled to match the endpoints of the desired operating temperature range by having the uCPU load the gain and offset registers with the proper binary value. The zero range is therefore settable from about -55 to +10 C and the full scale from +30 to +100 C without any external component change.

The output voltage which controls the oscillator frequency is produced by a 10-Bit charge integration D/A converter. The output is buffered and scaled to give an output range of 0 to +4.1 V.

Other analog functions included on the chip are: a 4.1 Vdc voltage reference and several gain stages for use as the crystal oscillator and buffer. A phase detector, reference counter and various analog switches are available to be used in the calibration mode.

This circuitry is integrated on a single die with dimensions of 5.77x7.20 mm (0.227x0.283 ") in a 1.5 micron process. The parts are presently packaged in a 28 pin PLCC although plans are underway to design a hybrid layout to use the ASIC in die form for further miniaturization.

Oscillator Calibration/Data Acquisition

In order to minimize the amount of testing time required and reduce the cost of building a test system, a unique calibration configuration was designed. An applied reference frequency is used to phase lock the crystal oscillator at its nominal frequency. The operation of the self-compensating procedure is described as follows:

1.) The initial set-up variables are loaded into the EEPROM through the serial data bus. These variables customize the internal program operation for the intended application by setting the clock frequency, loading the A/D gain and offset registers and adjusting the system timing.

2.) The unit to be calibrated is placed into a temperature chamber and the temperature is increased to the highest operating temp. After a stabilization period, a reference frequency from a synthesizer is applied to the unit. The uCPU is then initialized. The first task performed by the ROM program upon initialization is to interrogate the reference detection counter to determine if the reference signal is present. If so, it indicates to the program that a calibration run is being started.

In order to begin the calibration procedure, the uCPU places the phase locked loop section into the calibration configuration by asserting the calibrate mode select signal which closes an analog switch, closing the phase locked loop and phase locking the crystal oscillator to be compensated to the reference. It is, of course, necessary to have the center frequency and voltage tuning range of the crystal oscillator properly adjusted. A passive loop filter is connected between two external pins. As the ambient temperature subsequently changes, the tuning voltage will be adjusted by the loop to maintain phase lock as the crystal frequency tries to drift. Since the same voltage will be required to keep the oscillator at nominal frequency after the reference is removed, it can be digitized to determine what compensation data value is required.

When the calibration configuration is set, a switch is opened removing the negative feedback from the D/A amplifier so that it now functions as a high gain comparator instead of a unity gain buffer. Another switch is closed so that the uCPU may read the state of the comparator output from the D/A amp.

3.) After detecting the reference and setting the calibrate mode, the uCPU begins measuring the crystal temperature with the A/D converter. When the temperature has stabilized to within 1 Bit (approximately 0.04 Deg. C) over a period of time, a data point measurement is initiated. By controlling the D/A converter and using its buffer amp as a comparator, a successive approximation A/D conversion of the oscillator control voltage is performed. This determines to the nearest D/A bit what data will be required to achieve nominal frequency at this specific temperature when the reference is not present. This 10-Bit DAC frequency data and the corresponding 12-Bit temperature measurement define the x-y coordinates of a data point which is packed into a 3-Byte block and written into the EEPROM.

4.) The temperature is then lowered by the predetermined temperature increment and allowed to restabilize. Once again when the uCPU determines that the temperature has stabilized to within 1-Bit, a data point measurement is triggered and subsequently stored in the next sequential group of EEPROM locations.

5.) Step 4 is then repeated until reaching the lower limit of the operating temperature range. After the last data point has been acquired and stored, the chamber temperature is brought back up to room temperature. The uCPU senses this change, signifying the end of the calibration procedure. The data and status registers are then checked for any error flags which may have been set during calibration. A square wave signal is then output on one of the digital I/O pins, the frequency of which signifies the status of the run and the validity of the data.

6.) The reference frequency is then removed. This is sensed by the uCPU causing the program to branch to the normal frequency compensating mode of operation. The test system may then immediately begin a frequency verification run to determine the accuracy of the data.

Operational Mode

When the uCPU is initialized and the reference frequency is not present, the program immediately branches to the operate mode. It first of all makes sure that the configuration switches are set to the proper state, breaking the phase locked loop path and configuring the D/A amplifier as a unity gain buffer.

The program immediately measures the present temperature with the A/D. It then indexes through the data table looking for a stored temperature point close to the present temp. If a stored data point does not match the present temp. exactly, then the addresses above and below the present one are read into RAM along with their corresponding DAC data. A linear interpolation is then performed to determine the DAC voltage slope and to calculate the new DAC data required for the present temperature.

The typical temperature increment between the stored data points is 2 Deg. C. With about 40 bytes of the 256 byte EEPROM space used for storing set-up variables, this allows space for 72 data points to be stored. This would allow for a temperature span as wide as 144 Deg. C. with two degree steps. It has been shown that a look-up table with interpolation is capable of matching the crystal curve almost as well as a seventh order curve fit. (6) The linear interpolation was therefore chosen since it was much simpler and efficient to implement in machine code.

When the program is initialized, the uCPU continuously measures the temperature, interpolates the data and updates the DAC. If the A/D data shows little or no temperature change between readings, a number will be loaded into the standby timer register and the uCPU execution will be halted for some period of time. If the temperature is stable enough that there is still no significant delta, the standby time will be increased even further and eventually may extend for more than two minutes. This sleep time minimizes the DC power consumption and also shortens the amount of time that the uCPU will be generating digital noise.

When it is detected that the temperature is beginning to change, the standby time will be progressively shortened in order to track the temperature as closely as possible. By controlling the sleep mode duration, sampling interval, and the number of ADC samples per update cycle, a dynamic balance is maintained between the ability to track rapid temperature variations and the average power consumption.

Performance Results

Since most of the required components have been included in the ASIC chip design, only a crystal, a varactor diode and a few passive components are needed to implement a complete Digitally Compensated Crystal Oscillator. (See Figure 4)

After the initial calibration run has been completed, the data table may be read out of the EEPROM and examined if so desired. A graph of a typical table is shown in Figure 5. Since the compensation voltage is applied to the cathode of the varactor diode, a positive frequency vs. voltage transfer function results so that the data in

the table appears to be the inverse of the normal "AT" cut crystal curve.

Given a typical extended temperature range application of -40 to +85 Deg. C with a crystal peak-to-peak frequency deviation of 27 PPM, the maximum slope would be at the cold end of the crystal curve at about 1 PPM/Deg. C. Since the ASIC gives an A/D resolution of 12-Bits and a D/A resolution of 10-Bits, equation 2 on Figure 3 may then be used to calculate the best achievable stability in this application. Putting these numbers into the equation gives a temperature error term of 0.022 PPM and a voltage error term of 0.019 PPM which are summed together to give a total error of 0.041 PPM. Although this level may be achieved with great care, experience has shown that a factor of X2 to X3 should be added to allow for non-linearities and component anomalies which typically show up in production.

Figure 6 shows the frequency vs. temperature performance of this compensated oscillator on a scale of 0.05 PPM per division. This oscillator achieved a deviation of +/- 0.07 PPM compared to the predicted achievable value of 0.041 PPM.

Due to the small physical size of this system, it is inevitable that the CMOS uCPU and other digital circuitry will cause some degree of clock noise to appear on the oscillator signal. The chip however, was designed to segregate and minimize this feedthrough. The close to the carrier spurious signals are typically down -90 to -100 dBc and may be eliminated completely if the update inhibit line is asserted during critical periods. The oscillator board layouts have also been optimized to reduce the higher frequency clock feedthrough. Figure 8 shows a typical spectrum analyzer display of a 12 MHz oscillator with all clock noise close to -80 dBc.

Future Developments

As this project moves into production and the basic DCXO application begins to mature, work will continue on adding advanced functions. One area which is being evaluated is the possible implementation of long term aging correction to offset the drift of the crystal by changing the firmware to include a predetermined extrapolation algorithm.

This same chip may also be used for direct frequency compensation of DRO's, SAW's and other high frequency oscillators. It will also be possible to implement indirect frequency compensation using high-Q "SC" cut crystals for even better stability.

The variety of useful functions on the chip may be designed in various configurations by re-writing the firmware to perform accurate timing and counting, digital I/O, frequency synthesis, and various phase locking functions, possibly for integration with the end user's system.

Conclusion

This paper has described a practical ASIC which has been developed specifically as a single chip solution for temperature compensation of oscillator frequency. This versatile device should make possible many new applications such as re-compensation in the end user's equipment and remote frequency adjustment.

Acknowledgments

Several people deserve recognition for their work on this project; without their help, it would have been very difficult to pull this thing together:

>> Bob Antes, for writing thousands of lines of machine code and not missing a bit.
>> Tom Everingham, for the patience and perseverance to make things work like they're supposed to.
>> The Design Engineers at Sierra Semiconductor who proved that it could all fit on one chip.
>> The management of muRata who gave us the resources and the opportunity to make it happen.

References:

(1) G.E. Buroker, M.E. Frerking, "A Digitally Compensated TCXO", Proceedings of the 27th Annual Symposium on Frequency Control, June 1973, pp. 191-198

(2) G.B. Pollard, "A Digitally Compensated Hybrid Crystal Oscillator" Proceedings of the 36th Annual Symposium on Frequency Control, June 1982, pp. 474-485

(3) Marvin E. Frerking, "The Application of Microprocessors to Communications Equipment Design", Proceedings of the 33rd Annual Symposium on Frequency Control, May 1979, pp. 431-435

(4) Marvin E. Frerking, "Crystal Oscillator Design and Temperature Compensation", Van Nostrand Reinhold Co., New York, 1978, p. 155

(5) Steven Fry, "Digital Temperature Compensation for Oscillators", RF Design, February 1989, pp. 116-119

(6) A.M. Renard, K. Barnhill, "Digital Temperature Compensation of Crystal Oscillators", Proceedings of the 35th Annual Symposium on Frequency Control, May 1981, pp. 455-457

□

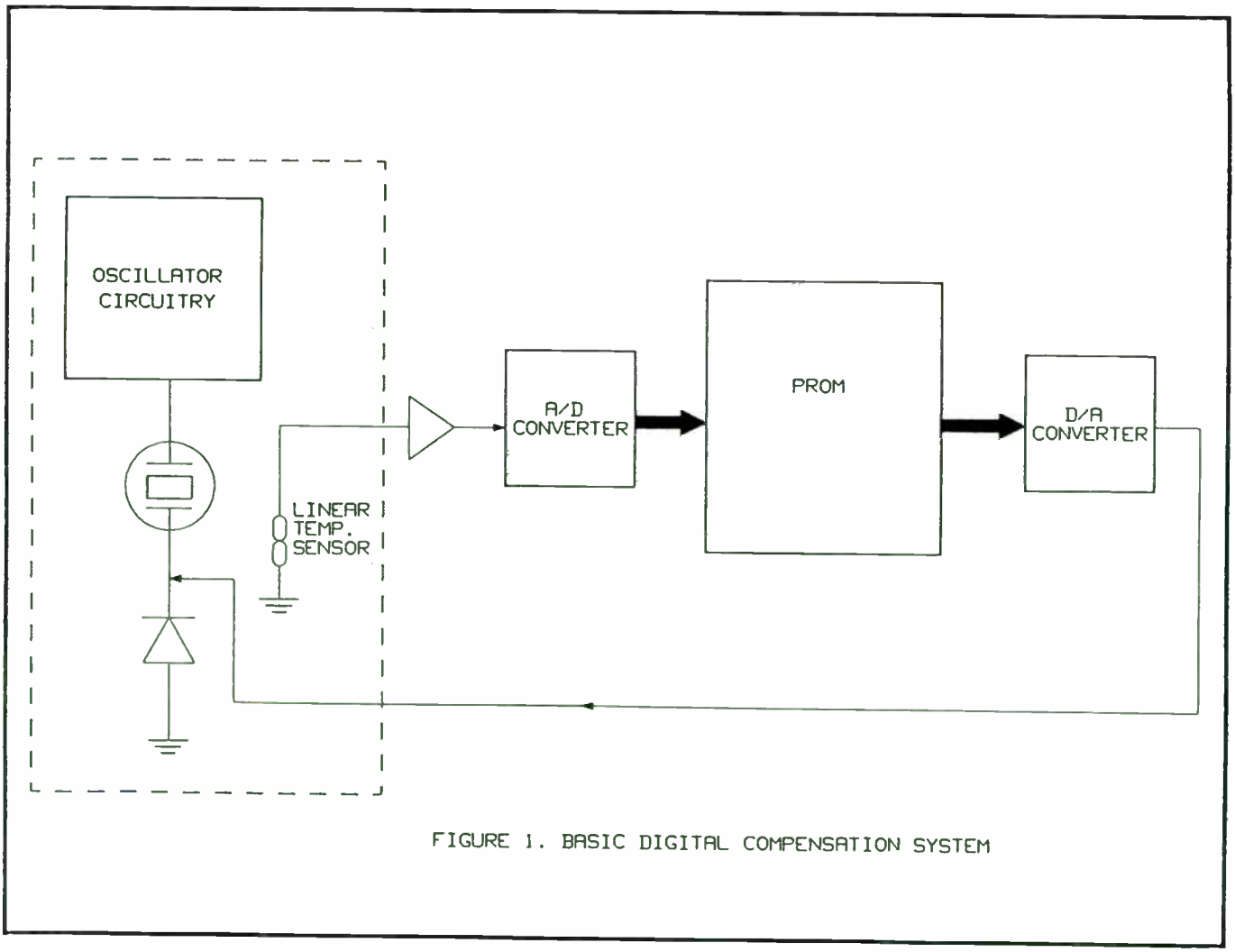


FIGURE 1. BASIC DIGITAL COMPENSATION SYSTEM

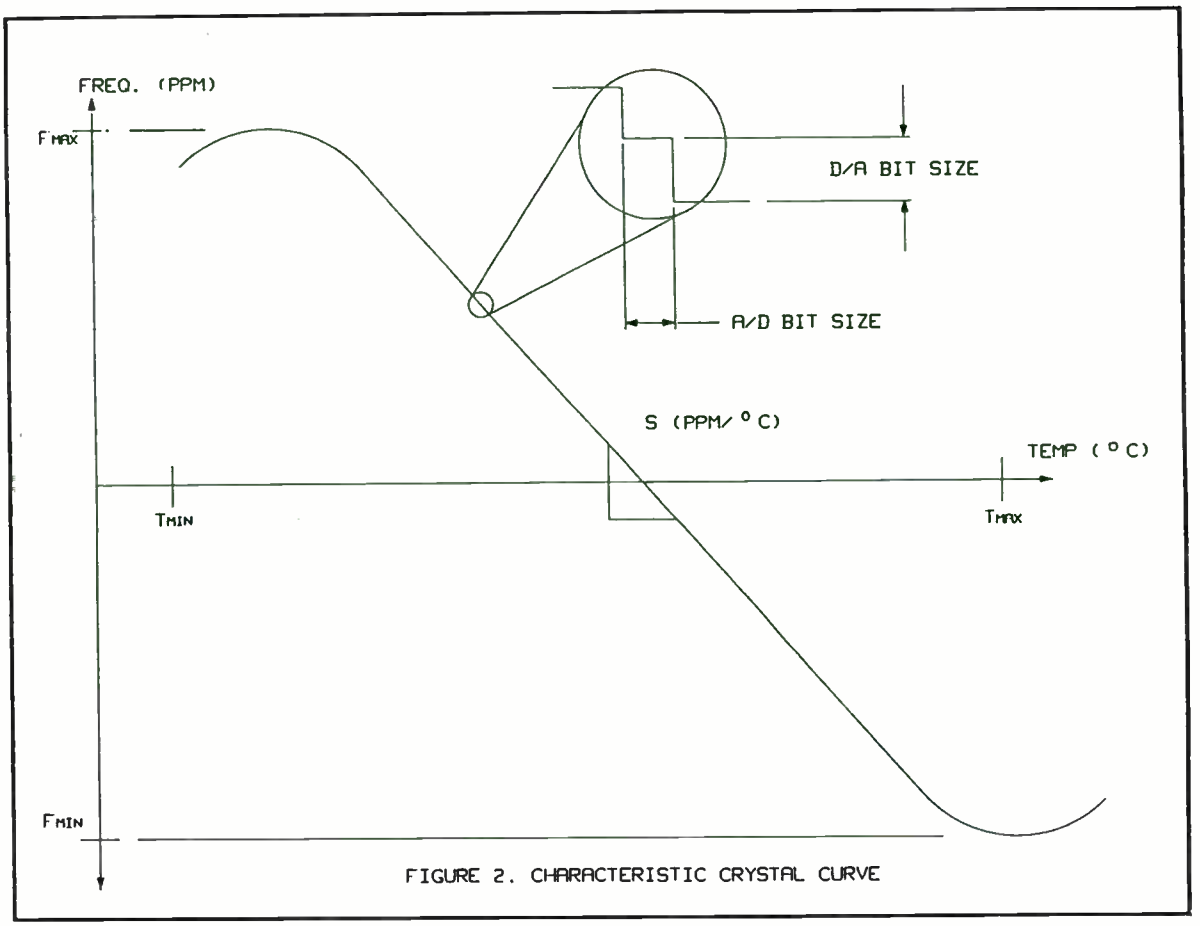


FIGURE 2. CHARACTERISTIC CRYSTAL CURVE

$$\Delta f = \left(\frac{S}{2}\right)\left(\frac{T}{2^{N_t}}\right) + \left(\frac{F}{2}\right)\left(\frac{1}{2^{N_v}}\right)$$

WHERE: $T = T_{MAX} - T_{MIN}$ (°C)

$F = F_{MAX} - F_{MIN}$ (PPM)

S = MAX. SLOPE OF CRYSTAL (PPM/°C)

N_t = # OF A/D BITS

N_v = # OF D/A BITS

ALLOWING FOR MANUFACTURING TOLERANCES:

$$\Delta f = \left(\frac{S}{2}\right)\left(\frac{T}{0.7\ 2^{N_t}}\right) + \left(\frac{F}{2}\right)\left(\frac{1}{0.7\ 2^{N_v}}\right)$$

FIGURE 3.

MAXIMUM ACHIEVABLE FREQUENCY STABILITY
OF DIGITAL COMPENSATION SYSTEMS

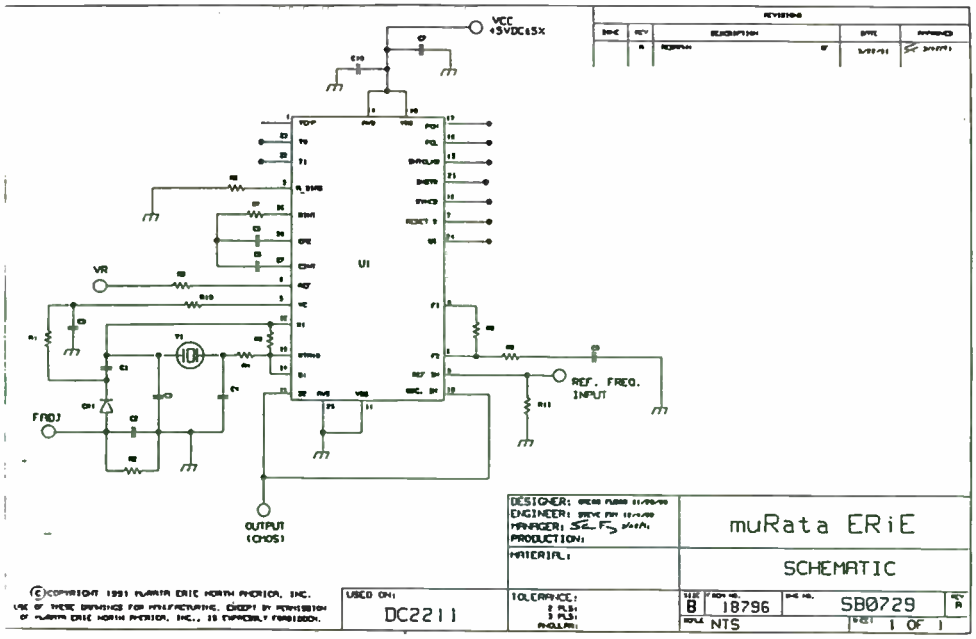


Fig. 4

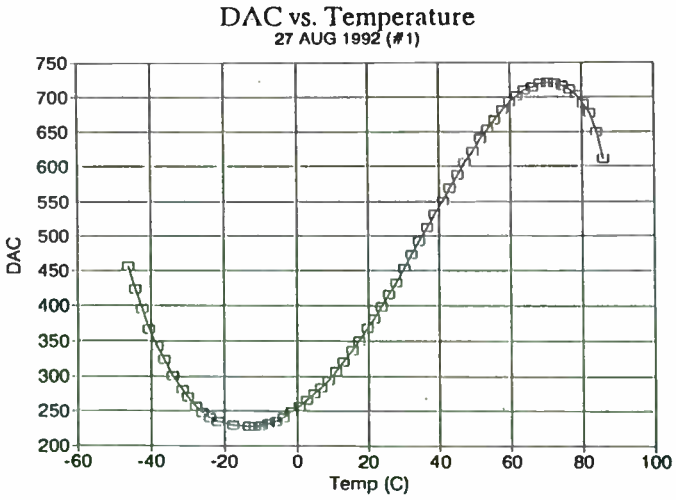


Fig. 5

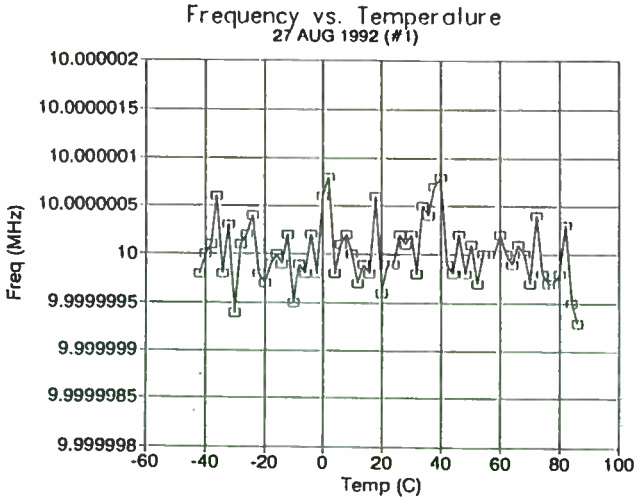


Fig. 6

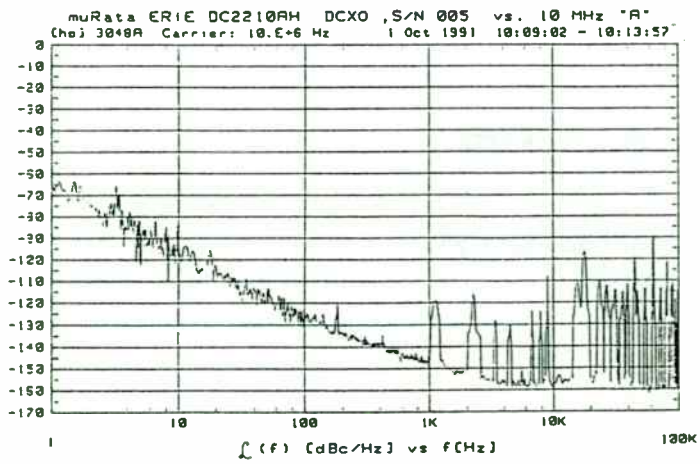


Fig. 7

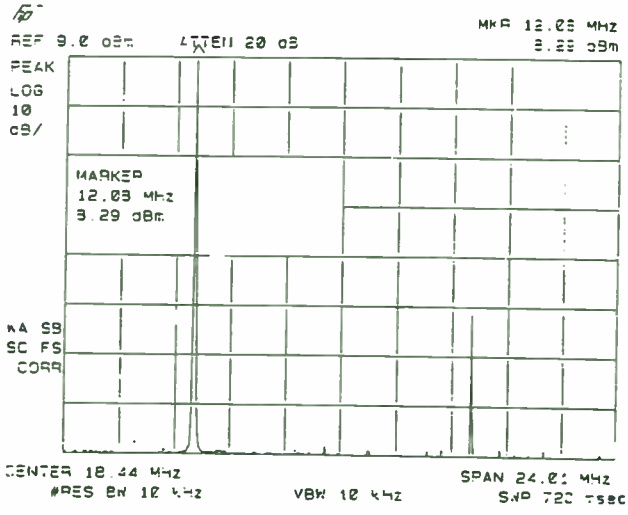


Fig. 8

A Synthesizer Design Program with Detailed Noise Analysis

Terry Hock
National Center for Atmospheric Research
Boulder, CO 80301

Abstract

There are several factors which determine or limit the phase noise of a synthesizer. The total phase noise of the synthesizer is determined by several noise sources in the loop such as the reference oscillator, voltage controlled oscillator (VCO), phase detector and operational amplifier (op amp) loop filter. Typically the op amp noise is considered a second or third order effect when a good low noise op amp is used, but in reality the op amp may be the limiting noise factor in a synthesizer. This program will analyze synthesizer taking into account all the noise sources in a synthesizer. The program will analyze and design the loop filter and calculate the loop steady state responses. All results are in both table and graphic form. It will also compute the reference spur levels for a digital phase/frequency detector with a differential output and provide the option to add an active low pass filter to reduce the spur level. The program was written to simplify the design of synthesizers using commercial synthesizer integrated circuits (IC). It operates on a DOS machine and requires a minimum of VGA graphics.

Overview

When designing a synthesizer, there are several performance criteria that must be considered: phase noise, switching speed, reference spur levels, etc. Usually optimizing one parameter gives less than adequate performance in another area. The performance of a single loop synthesizer is typically a compromise of all design goals.

In the past when designing synthesizers, I have used several design tools: spread sheets, small basic programs and spice simulators to analyze and design a synthesizer. This program was developed to combine all synthesizer design criteria into one easy to use program. A single program allows all aspects of a synthesizer design to be completed in a minimal amount of time. One parameter can be varied while the effect is observed in all performance areas of the synthesizer. Also in the past, an adequate model for the phase noise of the synthesizer was lacking in my design tools.

The program calculates the noise contribution from the individual stages and the total phase noise of the synthesizer. The loop bandwidth can easily be changed while observing the phase noise; this allows for easy optimization of the phase noise. The program displays the integrated phase jitter providing an aid in choosing the loop bandwidth for lowest overall phase noise. The loop filter model incorporates the performance of a real op amp for the calculations of op amp filter noise, reference spurs levels and steady state loop responses. The Johnson noise of the resistors is also incorporated into the op amp noise calculation.

All calculations use **linear control theory in the frequency domain** for both the steady state responses and phase noise calculations. A wideband synthesizer (loop bandwidth greater than 20% of the reference frequency) analyzed with linear analysis will introduce errors. A better approach would be to use discrete time analysis or Z-transforms for analysis. However, if wide loop bandwidths are desirable for fast switching speeds, the reference spurs may be at undesirable high levels when using a standard digital phase/frequency detector in a low cost synthesizer IC. To improve the accuracy of the calculations, the user has the option of adding a delay term to the steady state loop calculations. There have been considerable discussions in past issues of RF Design as to what the delay value should be [1]. The value of this delay term is dependent upon the type of phase detector (digital phase/frequency, sample & hold) used in the synthesizer and location of the first pole. There is not a clear consensus of what the sampling delay should be for all cases, so it's value is left to the user to choose.

Currently the program does not calculate the switching speed of the synthesizer, but uses a simple approximation for a rough estimate of the switching speed. Dan Gavin has demonstrated how a spice simulator can be used to estimate the switching speed

incorporating all delay terms for accurate switching speed analysis [2].

Phase Noise

When designing a synthesizer for lowest phase noise, optimum placement of the synthesizer loop bandwidth is critical. Typically only the VCO noise and phase detector noise floor are considered in determining the final phase noise of the synthesizer. Usually the phase noise inside the loop bandwidth is approximated by equation 1.

$$\text{phase noise} = 20 \log (N) + \text{phase det. noise floor} \quad (1)$$

N is the total division ratio of the VCO frequency to the reference frequency and the phase detector noise floor might be -155 dBc. The phase detector noise floor varies with the type of technology used such as CMOS, ECL and GaAs. CMOS has the lowest noise floor of the three. The loop bandwidth would be chosen where the noise floor from equation (1) intersects the phase noise of the VCO for optimum phase noise performance. This simplified analysis is shown in figure 1. As a quick evaluation of where to place the loop bandwidth, this procedure is quite valid. However this method excludes the true loop behavior or closed loop response and error loop response as they modify the VCO phase noise and the phase detector noise floor.

A more detailed analysis is required to predict the actual value of the phase noise in the synthesizer. As the architect of the synthesizer changes from large values of N to high VCO phase noise, this simple model will have significant errors in predicting the phase noise. A block diagram of all the noise sources that contribute to the phase noise is shown in figure 2. The op amp noise source is typically neglected when high quality low noise op amps are used in the loop filter integrator circuit. Yet in certain cases the op amp will be the major contributor to the phase noise, even when very low noise op amps are used, this is typically the case when a VCO with a high tuning sensitivity of 25 MHz/v or greater and phase detectors with low gain (1v/rad or less) are used.

Program Description

The program has four major sections: Design, Analysis, VCO phase noise and Synthesizer phase noise. The program assumes the user is familiar with designing synthesizers. Figure 3 is a schematic showing the circuit topology which the program evaluates. The synthesizer is a type II fifth order including the VCO bandwidth as a pole and first pole in a op amp. If an active filter is used to reduce the reference spurs, the synthesizer is a type II seventh order control system.

Design

The synthesizer design routine prompts the user for the common parameters in designing a synthesizer such as VCO tuning sensitivity (or gain), phase detector gain, reference frequency, loop bandwidth, etc. The component values of the loop filter
are calculated based on the equations from reference [3]. The program allows the user to easily change the component values for practical sizes in the loop filter without having to re-enter the initial design data. The new values are immediately updated on the screen. The program will next compute the value of the reference spurs at the VCO output. The spur calculation is only valid for a digital phase/frequency detector when using the differential phase detector outputs. This type of phase detector is used in the spur calculation since it has lower spur levels than a single ended output from a digital phase/frequency detector when the loop filter is an op amp integrator. The program then provides the option to add a 2-pole active low pass filter to reduce the spur level. The program prompts for bandwidth and damping coefficient of the filter. The filter component values are easily changed for practical values.

The design routine will also compute the steady state responses of the synthesizer both in table and graphic form. The steady state responses computed are closed loop gain, open loop gain, error loop gain and phase margin. The open loop gain and phase margin are the most critical parameters for evaluating the loop stability. The phase margin at 0 dB open loop gain frequency should topically be 40 to 60 degrees for a well behaved loop.

Usually, a synthesizer is designed to operate at several different frequencies. In a single loop synthesizer the output frequency may vary 25% or more, which requires changing the divider value thus the loop gain varies. The loop should be evaluated at several operating conditions for stability and bandwidth. Typically the tuning sensitivity of the VCO changes at different control voltages. This variation in the VCO gain must also be considered in a design. Both the tuning sensitivity and divider values can be evaluated quickly by entering either **K** or **F** followed by new values. The output frequency is changed in lieu of changing the actual divider ratio. The loop switching speed is also calculated based on a simple approximation of t=4/loop BW.

Analysis

The synthesizer analysis routine is very similar to the design routine except actual component values of the loop integrator and active low pass filter are entered. The steady state responses of the loop are calculated along with reference spur level. The steady state responses, open, closed and error loop are calculated in table and graphic form. The switching speed is not calculated since the analysis routine does not directly compute the loop bandwidth. The analysis routine is very valuable for analyzing current designs and to evaluate component sensitivity in the loop filter versus performance variations. The output frequency and the VCO gain can both be easily changed for different operating frequencies and VCO gain variations.

VCO Phase Noise

The VCO routine uses Leeson's equation for calculating the phase noise of an oscillator [4]. The two dominant parameters that set the phase noise are the loaded Q of the resonator and the operating frequency or the ratio of operating frequency to Q. The output power and flicker frequency are also required in the VCO phase noise calculation. The varacter diode is also a major noise source in a voltage controlled oscillator. The varacter diode has an effective thermal noise resistance. This resistance generates a noise voltage given by Nyquist's equation which modulates the varacter diode. This noise source typically dominates the phase noise of the VCO in wide frequency tuning oscillators.

The VCO routine prompts the user for operating frequency, loaded resonator Q, output power, flicker frequency, effective noise resistance of varacter diode and tuning sensitivity. The VCO phase noise analysis separately calculates the phase noise due to the varacter diode, Leeson's equation and the total phase noise. The performance of the VCO is easily evaluated showing which factors dominated the VCO phase noise. The results are in table and graphic form.

The resonator Q and tuning sensitivity can easily be changed in 5% increments up or down. This feature is useful for adjusting the phase noise performance to match that of a commercial VCO that may be used in the synthesizer design. When the Q and tuning sensitivity are changed, the results are updated real time on the screen. **The phase noise from this routine will be used in the synthesizer phase noise calculations.** When designing VCO's this routine is very useful for estimating the phase noise performance of an oscillator.

Synthesizer Phase Noise

The synthesizer phase noise routine is the most useful feature of the program. It computes the individual phase noise terms of each noise source in the loop and simultaneously displays the level of the reference spur, switching speed and integrated phase jitter. All design parameters are displayed on one screen while the user can easily change the loop bandwidth while observing the phase noise, spur level, switching speed, etc. The output frequency and VCO tuning sensitivity can also be changed to account for the different operating conditions as the synthesizer is set to different frequencies.

The synthesizer phase noise calculation uses data from either the Design of Analysis routine for the loop filter and data from the VCO phase noise routine. Phase noise is computed for the op amp filter, reference oscillator, VCO, phase detector and the sum of these noise sources. The op amp noise model also includes the active low pass filter if it is chosen in either the analysis or design routines. The model includes the thermal noise of all the resistors in the filters. The phase jitter is also computed by integrating the total phase noise from 100 Hz to 1 MHz. The phase jitter is useful when adjusting the loop bandwidth for lowest phase noise.

The loop bandwidth, tuning sensitivity and output frequency can all be easily changed in either the table or graphic display of the synthesizer phase noise. When a change is made, all data is updated automatically on the screen. When the loop bandwidth is changed, the feedback capacitor is held constant so all resistor values change. This is very important when the loop bandwidth is reduced as all the resistor values will increase in thus increasing the op amp noise level. The loop bandwidth is not an option for changing its value when using data from the analysis routine. The phase noise plot has five parameters plotted simultaneously, so when a parameter like VCO gain is changed several times the graph becomes cumbersome to evaluate. A redraw feature is available that will update the display with the last data set by simply pressing **R**.

Miscellaneous Features

The program also has a help file for operating the program, but it is assumed that the user is familiar with synthesizer design. There are also a few common parameters such as sampling delay and resistance tolerance that seldom require changing; they are entered in a special utility routine. The resistance tolerance is

required for the spur calculation. The spur calculation computes the common mode rejection of the op amp integrator from the op amp data and resistor values. The op amp data is in a separate ASCII file that contains the individual parameters for each op amp. This ASCII data file can easily be changed with any text editor to add, delete, or modify the op amp data file. The read.me file provides the format for the ASCII op amp data file. A second ASCII data file is required which stores all synthesizer data parameters as the default values when the program is terminated. These parameters will then be recalled when the program is run again. The op amp and PLL data files must reside in the same directory when the program is executed for proper operation.

General Operation

The program is designed to be very easy to use and operate with a minimal number of key strokes. At each prompt there will be a question which will ask for a value, such as phase detector gain, or a single letter for an option. At the end of a prompt, a number will be in brackets which is the default value. By simply pressing return, the program excepts the default value. When only one parameter has to be changed, this feature quickly allows one to go through the input parameters without having to retype in every value. An option selection will have different letters in brackets; pressing the letter will execute that command.

Printing

The results on the screen can easily be dumped to a printer using the PrintScreen key. A printed hard copy of the graphs can also be dumped to a printer by running the DOS utility GRAPHICS.COM prior to running the program. To dump the graphs to a printer, simply press [Shift] and [Print Screen] simultaneously. If a Laser printer is being used, be sure to add the appropriate extension to the graphics command given in a DOS manual.

Phase Noise Model

Several assumptions are made in the computation of the phase noise. The steady state equations used for the analysis and design routine are used for the phase noise calculations in modifying the appropriate noise source in the loop. The Z-transforms would be the better choice for improved accuracy in wide loop bandwidth designed synthesizers. However for small loop bandwidths relative to the reference frequency, using linear equations provides excellent results. The most unique feature of the program is that it incorporates all noise contributions from the loop's integrator and active low pass filter. In designing low phase noise synthesizers the op amp noise can limit the phase noise performance.

A block diagram of the noise sources considered in this program are shown in figure 4. The noise sources E1(s) and E2(s) are lumped together into the op amp noise source when displayed on the screen. E1(s) represents the output noise of the integrator circuit including the thermal noise of all the resistors. E2(s) represents the output noise of the active low pass filter including the thermal noise of the resistors. The noise from the low pass filter is only considered if selected in the design or analysis routines. Since most synthesizers use CMOS technology, the final frequency division of the VCO is with CMOS technology. CMOS has the lowest phase noise of all technologies available. Therefore the frequency divider phase noise was omitted from the analysis. However, the reference oscillator noise floor and divider noise are equivalent in the system equations. The phase noise of the reference oscillator is based upon Leeson's equation with a very high Q representative of a crystal oscillator. The reference oscillator usually dominates the total phase noise below 100 Hz in designs with high divider values.

An accurate model of the phase detector noise floor is not available for the various types of technologies and types of phase detectors so a constant value is entered by the user. Even though the phase noise density of the phase detector is not constant versus frequency, the noise floor will dominate the total phase noise typically at only one frequency range. The value entered should correlate to this frequency region.

The noise model of the loop filter accounts for the internal noise generated by the op amps (input noise current density and input noise voltage density) and the thermal noise of the resistors. Figure 4 shows the noise sources considered in the filter noise model used by the program. Only the Johnson or thermal noise of the resistors is considered. Contact or popcorn noise in the resistor is not considered, as this noise is very dependent on the type of resistor used and is not well defined. Metal film resistors are a good choice in the integrator filter for low noise since they have minimal contact noise. The output noise of each filter stage is computed for the calculation of the phase noise. The flicker frequency noise of the op amps, typically DC to 100 Hz is not considered in the calculation of the op amp noise. In many cases the data sheets for op amps do not provide complete noise information to consider the low

frequency flicker noise. Therefore this noise source is not considered in the model. The phase noise of the synthesizer is typically dominated by the reference oscillator at these low frequencies, as a result the addition of the flicker noise in the op amp would add little more accuracy in the calculations.

The op amp noise is minimized by keeping the input resistors of the integrator circuit small. The input noise current is effectively magnified by the size of the input resistors. Also the Johnson noise from the input resistors can become excessive if the resistors are large. If the low pass active filter is implemented, it does contribute noise to the overall loop filter and is added to the op amp noise display.

The program requests a VCO bandwidth which is used in the steady state analysis and synthesizer phase noise calculations. This is an additional pole in the loop. The VCO bandwidth in the analysis is represented as a single RC low pass filter. The VCO bandwidth can be either the actual bandwidth of the VCO or an RC low pass filter located at the input of the VCO. This additional pole filters the noise generated in the op amp circuits which can improve the phase noise performance when op amp noise is high.

Example

It is very easy to evaluate which is the dominating noise source at a given offset frequency from the carrier in a synthesizer with this program. In loops with high division values, the phase noise inside the loop bandwidth is typically dominated by the phase detector noise floor and the loop filter (op amp noise). If a VCO with poor phase noise is used, it still will contribute significant noise inside the loop bandwidth as there is not enough loop gain to reduce the VCO noise. The phase noise is dominated again by the op amp noise and VCO noise outside the loop bandwidth. It becomes clear that for low phase detector gains and large VCO tuning sensitivities the loop filter is a major noise contribution to the overall performance.

The op amp noise can be minimized by using a VCO which has a lower tuning sensitivity or a phase detector with higher gain. A sample and hold phase detector with high gain of 10 v/rad or more may significantly improve the overall phase noise my minimizing the op amp noise. A second approach would be to use a phase detector whose output is a current driver that does not require an op amp yet is still a type II control system [5].

A synthesizer operating at 1280 MHz with a reference frequency of 250 kHz was designed and analyzed using this program. The main design goal was for low phase noise using a commercial VCO, good sideband reference suppression greater than 60 dB, and switching speeds less than 5 msec. Figure 5 is a plot of the measured phase noise of the synthesizer and figure 6 is a plot of the calculated data from the program. Note the small noise peak at an offset of 3 kHz in the measured data, this agrees with the predicted from the program. The noise peak may be mistaken for a marginally stable loop with a low damping coefficient but the frequency step response show that the loop is well behaved. There is quite good agreement between measured data and calculated data. The measured switching speed for a 200 MHz change is shown in figure 7. The simple switching speed approximation in the program does give a good first order approximation when compared to measured results.

Conclusion

It is very easy to evaluate which component in a synthesizer is the dominating noise source at a given offset frequency from the carrier with this program. The engineer can quickly evaluate several different synthesizer architects for best performance without having to prototype a synthesizer. In today's competitive market, an accurate synthesizer model is extremely valuable in minimizing development time of a new synthesizer. The program has been shown to have good agreement with measured data of an actual synthesizer. I would especially like to thank Mitch Randall and Tom Thompson for their valuable feedback in evaluating the program as it was developed.

References

1. Egan, William, Sampling Delay - Is It Real?, RF Design, Cardiff Publishing, February 1991.

2. Gavin, Dan T., Phase Locked Loop Synthesizer Switching Speed Optimization and Measurement Techniques, RF EXPO Proceedings West 1989, Santa Clara, CA pp. 355-372.

3. Przedpelski, Andrzej B., PLL Primer Part III, RF Design, Cardiff Publishing, July/August 1983.

4. Robins, W. P., PHASE NOISE IN SIGNAL SOURCES THEORY AND APPLICATIONS, Peter Peregrinus Ltd., 1982.

5. Beam, William C. and Rezin, Phillip J., Design

Considerations for Fast Switching PLL Synthesizers, Watkins-Johnson Co. Tech Note, VOL. 18 NO. 2 March/April 1991.

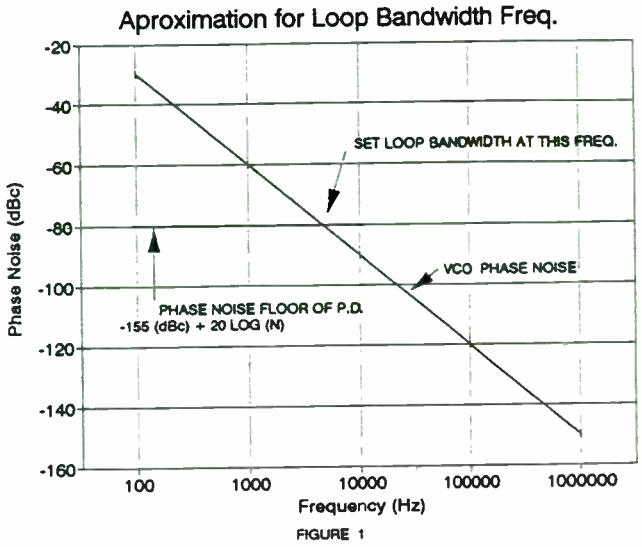


FIGURE 1

PHASE NOISE SYSTEM EQUATIONS

$$\text{REF OSC PHASE NOISE} = \frac{\frac{Kpd\ Kvco\ F_1(s)\ F_2(s)\ F_3(s)}{s}}{1 + \frac{Kvco\ Kpd\ F_1(s)\ F_2(s)\ F_3(s)}{s\ N}} \cdot \theta_{Ref}(s)$$

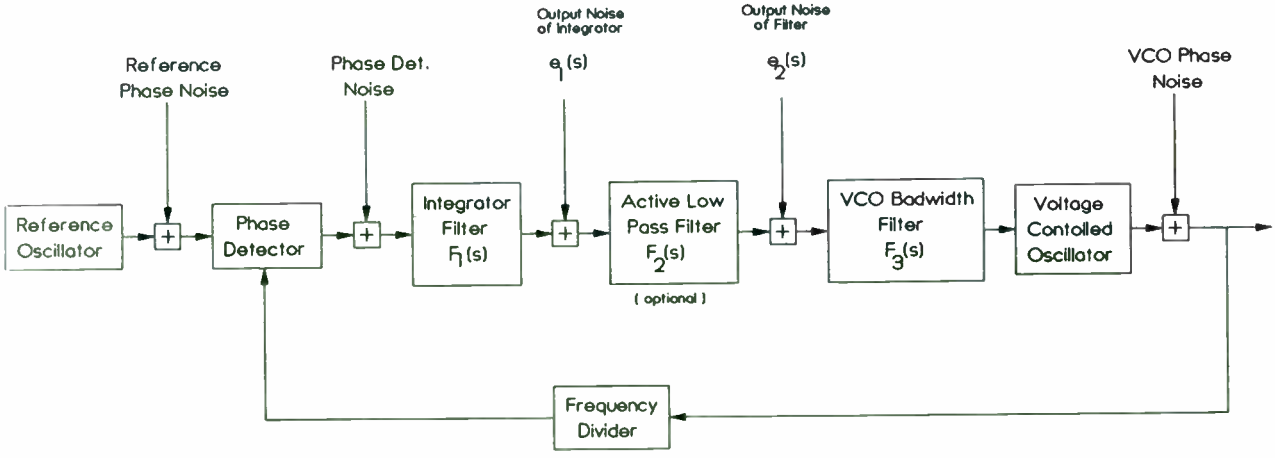
$$\text{PHASE DET. PHASE NOISE} = \frac{\frac{Kvco\ F_1(s)\ F_2(s)\ F_3(s)}{s}}{1 + \frac{Kvco\ Kpd\ F_1(s)\ F_2(s)\ F_3(s)}{s\ N}} \cdot \theta_{PD}(s)$$

$$\text{VCO PHASE NOISE} = \frac{1}{1 + \frac{Kvco\ Kpd\ F_1(s)\ F_2(s)\ F_3(s)}{s\ N}} \cdot \theta_{VCO}(s)$$

$$\text{OP-AMP INTEGRATOR NOISE} = \frac{\frac{Kvco\ F_2(s)\ F_3(s)}{s}}{1 + \frac{Kvco\ Kpd\ F_1(s)\ F_2(s)\ F_3(s)}{s\ N}} \cdot E_1(s)$$

$$\text{OP-AMP LOW PASS FILTER NOISE} = \frac{\frac{Kvco\ F_3(s)}{s}}{1 + \frac{Kvco\ Kpd\ F_1(s)\ F_2(s)\ F_3(s)}{s\ N}} \cdot E_2(s)$$

FIGURE 2A



Noise Sources in Synthesizer Figure 2

SYNTHESIZER TOPOLOGY

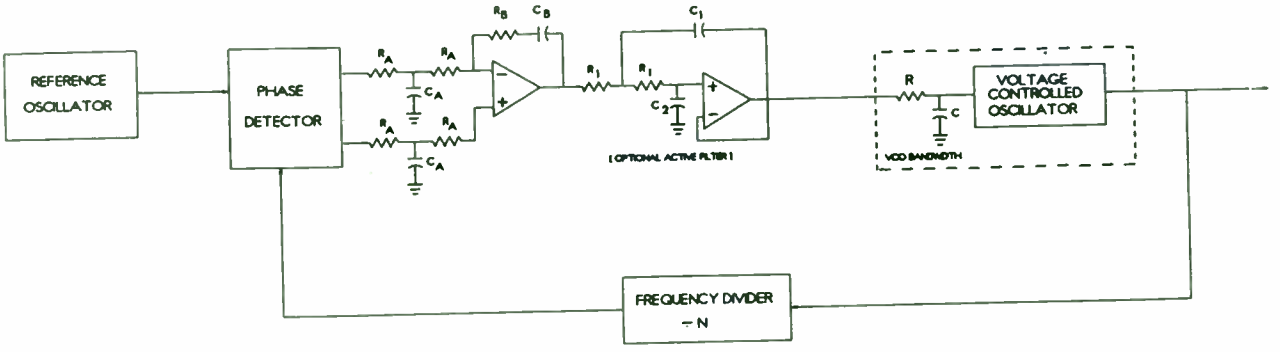


FIGURE 3

LOOP FILTER NOISE MODEL

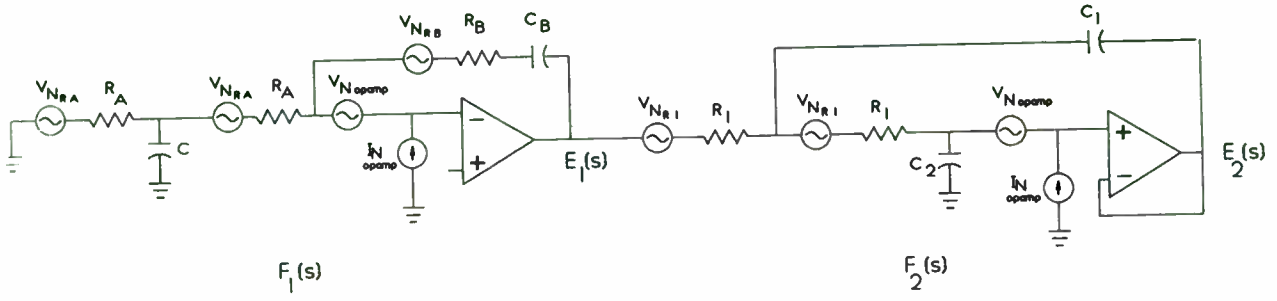
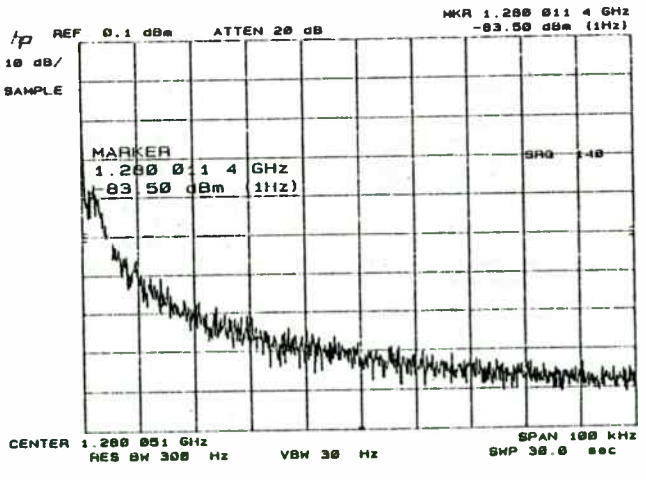
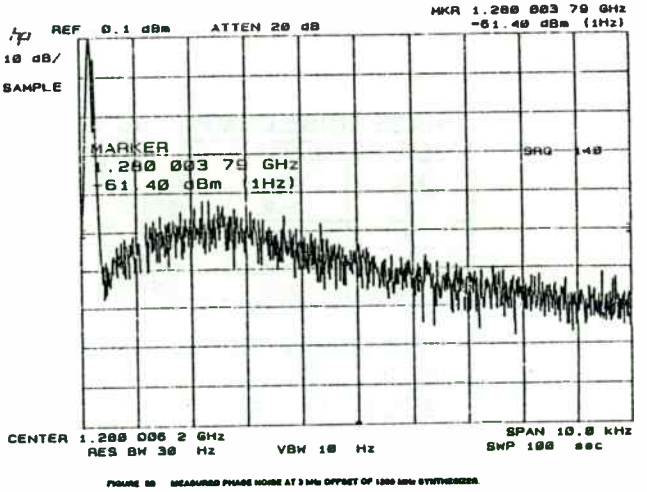


FIGURE 4





Frequency (Hz)	FILTER (dBc)	VCO (dBc)	REF (dBc)	Phase Det. (dBc)	Total (dBc)	
10.0	-80.1	-84.3	-64.7	-80.0	-64.4	
17.8	-80.1	-81.0	-69.6	-80.0	-68.7	Fvco: 1280.00 MHz
31.6	-80.1	-79.3	-74.4	-80.0	-71.8	
56.2	-80.1	-76.0	-79.0	-80.0	-72.9	Fref: 250.00 kHz
100.0	-80.0	-74.3	-83.4	-80.0	-72.2	
177.8	-79.8	-71.0	-87.3	-80.7	-70.6	Kvco: 23.0 MHz/v
316.2	-79.1	-69.5	-90.6	-80.5	-68.7	
562.3	-77.7	-67.3	-92.9	-79.9	-66.7	Kpd: 0.796 v/rad
1000.0	-75.6	-65.6	-93.9	-78.0	-65.0	
1778.3	-73.4	-64.0	-94.0	-77.4	-64.1	
3162.3	-73.5	-66.6	-95.6	-78.1	-65.6	
5623.4	-80.6	-75.1	-104.3	-86.2	-73.0	
10000.0	-89.0	-84.0	-116.1	-97.6	-83.4	Phase jitter over
17782.0	-98.0	-92.9	-128.9	-110.2	-91.0	100 Hz to 1 MHz
31622.0	-108.1	-108.1	-142.0	-124.0	-99.5	2.097 deg.
56234.1	-117.0	-106.9	-157.4	-138.5	-106.5	
100000.0	-127.7	-113.3	-172.3	-153.4	-113.1	
177827.9	-137.7	-119.3	-187.3	-168.4	-119.2	
316227.0	-147.7	-125.0	-202.6	-183.7	-125.0	
562341.3	-157.7	-130.5	-218.4	-199.5	-130.5	
1000000.0	-167.7	-135.0	-235.2	-216.2	-135.0	

FIGURE 6A CALCULATED PHASE NOISE RESULTS.

```
Input Parameters:
-------------------------------------------
Output Frequency:      1280 MHz
Reference Frequency:   250 kHz
Phase Detector Gain:   .796 V/rads
VCO Gain:              23 MHz/V
Total division N:      5120
VCO Bandwidth:         8 kHz
Op Amp:                LT1812
Resistance Tolerance:  5 %
Ra: 2.43 kohms
Rb: 3.74 kohms
Ca: 12 nF
Cb: 47 nF
```

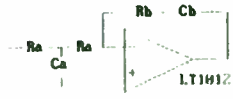


FIGURE 6B PARAMENTERS USED FOR CALCULATION OF 1280 MHz SYNTHESIZER PHASE NOISE.

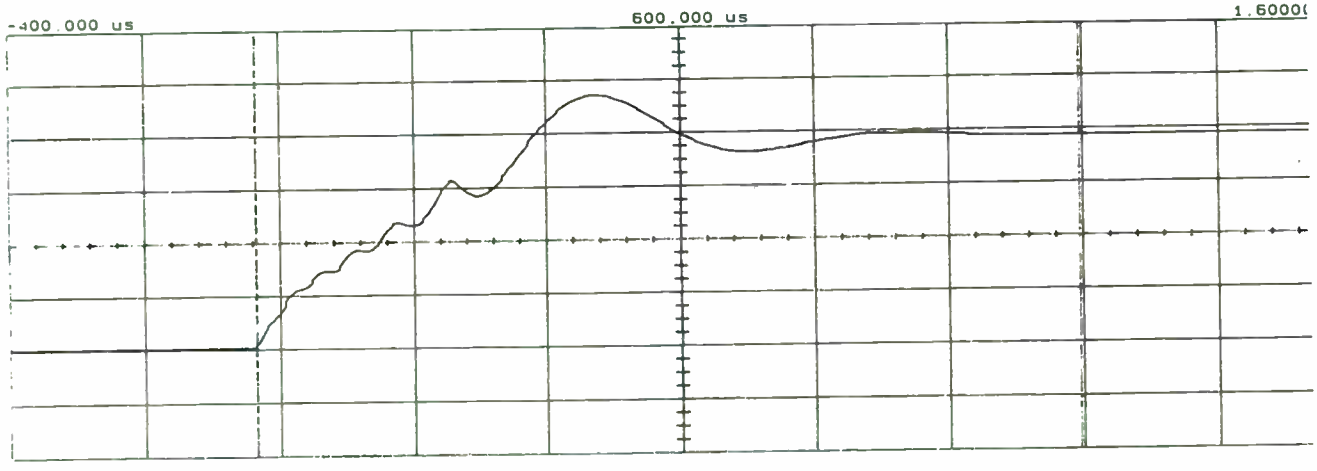


FIGURE 8 MEASURED SWITCHING SPEED OF 1280 MHz SYNTHESIZER FOR 200 MHz FREQUENCY STEP.

PHASE NOISE COMPARISON

FREQUENCY kHz	MEASURED dBc	CALCULATE dBc
1	-66.9	-65.0
3	-61.4	-65.6
10	-83.5	-83.4
50	-106	-106.0

SWITCHING SPEED

MEASURED	1.2 msec
CALCUALTE	1.54 msec

FIGURE 7

```
Input Parameters:                    Calculated Filter Parameters
-------------------------------------------------------------------
Output Frequency:      480 MHz       Ra =  10.870 k ohms
Reference Frequency:   50 kHz        Rb =  57.820 k ohms
Phase Detector Gain:   .796 V/rads   Ca =   7.11 nF
VCO Gain:              5 MHz/V       Cb =   6.800 nF
Total division N:      9600
VCO Bandwidth:         15 kHz        R1 =   5.684 k ohms
Op Amp:                LF156         C1 =   5.60 nF
Resistance Tolerance:  5 %           C2 =   1.400 nF
Loop Bandwidth:        1.3 kHz
Phase Margin:          50 deg.

Low Pass BW 3dB:       10 kHz
Low Pass Damping:      .5
```

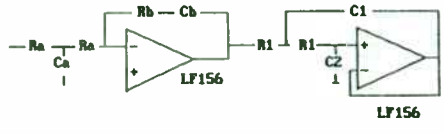


FIGURE 9 DISPLAY OF INPUT DESIGN SCREEN.

FREQUENCY (Hz)	OPEN LOOP GAIN Gain (dB)	OPEN LOOP GAIN Angle (deg)	CLOSED LOOP Gain (dB)	ERROR LOOP Gain (dB)	50.0 kHz REFERENCE SPUR ANALYSIS
10.0	74.6	1.2	78.1	-74.6	
17.8	64.6	2.1	78.1	-64.6	Spur level with only Op-Amp
31.6	54.6	3.7	78.1	-54.6	integrator (Ra,Rb,Ca,Cb).
56.2	44.7	6.5	78.1	-44.6	INITIAL LEVEL: -44.2 dBc
100.0	34.8	11.3	78.2	-34.7	
177.8	25.3	19.2	78.5	-24.9	
316.2	16.6	30.2	79.2	-15.4	Attenuation Phase
562.3	9.1	48.7	80.3	-6.8	0 50.0kHz 0 1.3kHz
1000.0	2.8	64.4	81.2	0.3	(dB) (deg)
1778.3	-3.1	36.4	79.3	4.4	
3162.3	-9.4	13.7	72.0	3.4	VCO BW: 10.8 -5.0
5623.4	-16.8	-28.1	62.5	1.2	ACTIVE LPF: 27.8 -7.5
10000.0	-27.7	-103.8	50.3	-0.1	
17782.8	-40.4	-178.9	29.6	-0.8	TOTAL 38.6 -12.5
31622.8	-72.4	141.0	5.6	-0.8	
56234.1	-97.1	118.5	-19.0	-0.8	Final Reference Spur Level:
100000.0	-122.8	105.3	-43.9	-0.8	-82.8 dBc @ 50.0 kHz
177827.9	-146.9	97.2	-68.9	0.0	
316227.8	-171.9	91.6	-93.9	0.0	
562341.3	-196.9	86.5	-118.9	0.0	Aproximate switching speed:
1000000.0	-222.1	80.4	-144.0	0.0	3.00 msec

FIGURE 10 DISPLAY OF STEADY STATE RESPONSES, SPUR AND SWITCHING SPEED CALCULATION.

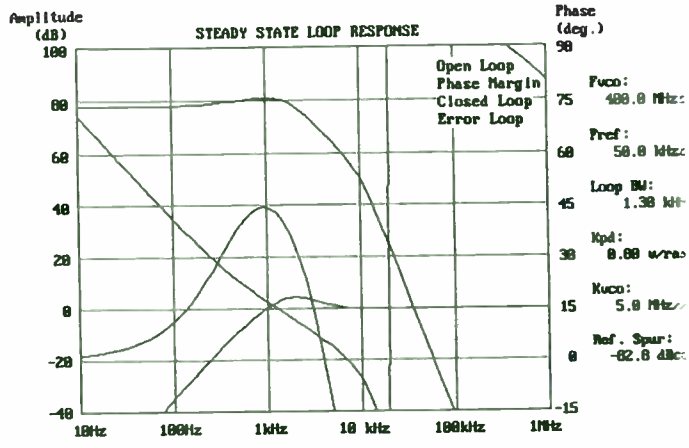


FIGURE 11 GRAPH OF STEADY STATE RESPONSES.

Frequency (Hz)	Loaded Q (dBc)	Varac. Diode (dBc)	TOTAL (dBc)	Input Parameters:
10	-19.5	-49.9	-19.5	VCO frequency: 400 MHz
18	-27.0	-54.9	-27.0	VCO loaded Q: 6.0
32	-34.5	-59.9	-34.5	Power output: 10 dBm
56	-42.0	-64.9	-42.0	Flicker frequency: 5000 Hz
100	-49.5	-69.9	-49.4	VCO tuning sensitivity: 5.0 MHz/v
178	-56.9	-74.9	-56.8	Varacter noise resistance: 10.0 kohms
316	-64.3	-79.9	-64.2	
562	-71.6	-84.9	-71.4	RMS phase jitter: 100 Hz to 1 MHz:
1000	-78.8	-89.9	-78.4	1.966 deg.
1778	-85.7	-94.9	-85.2	
3162	-92.4	-99.9	-91.7	Press:
5623	-98.8	-104.9	-97.8	[1] Increase Q
10000	-104.8	-109.9	-103.6	[2] decrease Q
17783	-110.5	-114.9	-109.1	[3] Increase Kvco
31623	-115.9	-119.9	-114.4	[4] decrease Kvco
56234	-121.2	-124.9	-119.6	
100000	-126.3	-129.9	-124.7	
177828	-131.4	-134.9	-129.8	
316228	-136.5	-139.9	-134.8	
562341	-141.5	-144.9	-139.8	
1000000	-146.5	-149.9	-144.9	

FIGURE 12 DISPLAY OF VCO PERFORMANCE.

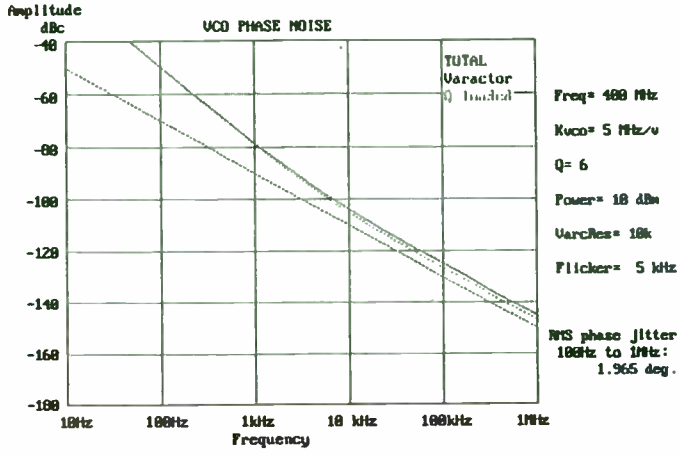


FIGURE 13 GRAPH OF VCO PERFORMANCE.

Frequency (Hz)	FILTER (dBc)	VCO (dBc)	REF (dBc)	Phase Det. (dBc)	Total (dBc)	
10.0	-78.9	-94.1	-68.8	-76.9	-68.7	
17.8	-78.8	-91.6	-65.7	-76.9	-65.2	Fvco: 400.00 MHz
31.6	-78.7	-89.1	-78.5	-76.9	-69.1	
56.2	-78.5	-86.6	-75.1	-76.9	-71.7	Pref: 50.00 kHz
100.0	-78.1	-84.1	-79.4	-76.8	-72.8	
177.8	-77.1	-81.7	-83.1	-76.5	-72.7	Kvco: 5.0 MHz/v
316.2	-75.5	-79.6	-85.9	-75.8	-71.7	
562.3	-73.5	-78.2	-87.6	-74.7	-70.2	Kpd: 0.796 v/rad
1000.0	-71.8	-78.1	-88.9	-73.8	-69.1	
1778.3	-72.6	-80.8	-92.3	-75.7	-70.4	Loop Bandwidth,
3162.3	-77.8	-88.3	-100.5	-83.0	-76.3	Fo: 1.3 kHz
5623.4	-83.8	-96.6	-110.7	-92.5	-83.1	
10000.0	-91.0	-103.7	-123.2	-104.7	-90.6	Phase jitter over
17782.8	-105.6	-109.1	-144.0	-125.4	-104.8	100 Hz to 1 MHz
31622.8	-121.2	-114.4	-160.2	-149.4	-113.6	1.322 deg.
56234.1	-134.6	-119.6	-192.9	-174.0	-119.5	
100000.0	-146.1	-124.7	-217.8	-198.9	-124.7	
177827.9	-156.7	-129.8	-242.8	-223.9	-129.8	
316227.8	-166.9	-134.8	-267.8	-248.9	-134.8	
562341.3	-176.9	-139.8	-292.8	-273.9	-139.8	
1000000.0	-186.9	-144.9	-317.9	-299.0	-144.9	

FIGURE 14 DISPLAY OF SYNTHESIZER PHASE NOISE.

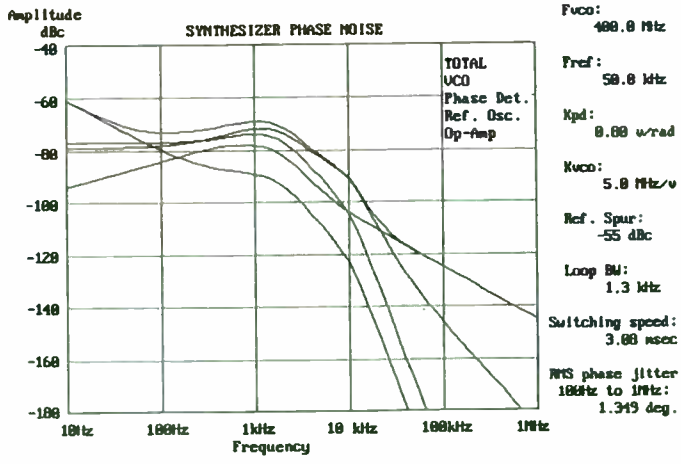


FIGURE 15 GRAPH OF SYNTHESIZER PHASE NOISE.

PLL Settling Time: Phase vs Frequency

Donald E. Phillips
Rockwell International
400 Collins Road
Cedar Rapids, IA 52498

Frequency synthesizers for frequency hopping systems need to tune rapidly to permit efficient transmission of data during each hop period. The tuning speed is usually specified by either frequency or phase settling time. The purpose of this paper is to show that the phase settling characteristics correspond to system performance better than the frequency settling characteristics. A simulation will be used to clarify this paradoxical relationship intuitively, followed by a mathematical analysis.

There is a performance loss in a frequency hopping FSK system where each hop period is degraded by the time used to tune the synthesizer to each new frequency. For systems that are not phase-coherent (where the phase of each hop is independent of the phase of other hops) the signal during each hop is integrated over the hop period to obtain the best signal to noise ratio. Various techniques are used such as matched filters, FFT or other digital filters, etc. The following simulation uses a high-Q filter to integrate the signal over the first part of a hop where the synthesizer settling occurs. The results will be compared to an analysis using pure integration.

Simulation

The simulation circuit of figure 1 consists of a variable frequency source driving the two inputs to identical band-pass filters. These inputs are 90 degrees apart in phase. The filters are series resonant circuits L, C, and R. The output voltages across the resistors are squared and summed to indicate the power out of the filter. One circuit would give the same result but with a more uneven power output and a poorer plot.

This band-pass filter (BPF) is normalized to a 1 rad/sec resonance. Since R = 1 ohm, the filter Q is set by the L and C values. For this example, a Q of 50 resulted in C = 1 / Q = 0.02, and L = Q = 50.

The four state-variables are the inductor currents and capacitor voltages, which are found by a differential equation solver. Two solutions were found, one which started on the desired frequency (1 rad/sec) centered in the pass-band at t = 0. The other started with a considerably higher input frequency of 5 rad/sec and settled to the desired frequency at an exponential rate simulating a phase-locked loop.

Since the filter Q = 50, the bandwidth is 0.02, or from 0.99 to 1.01 rad/sec. The frequency change compared to one side of the passband is (5 - 1) / 0.01 = 400.

The phase-locked loop bandwidth is 0.2 rad/sec, a time-constant t of 5 seconds. A first-order loop equation is used since higher order loops end up with the same settling pattern. The early part of the transient consists of many phase oscillations from which no useful output is obtained. The phase error eventually reaches a value of zero, arbitrarily assigned as the final phase value.

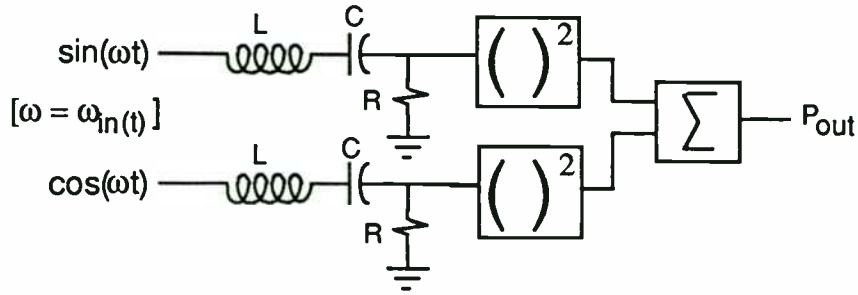


Fig. 1 - Simulation Model

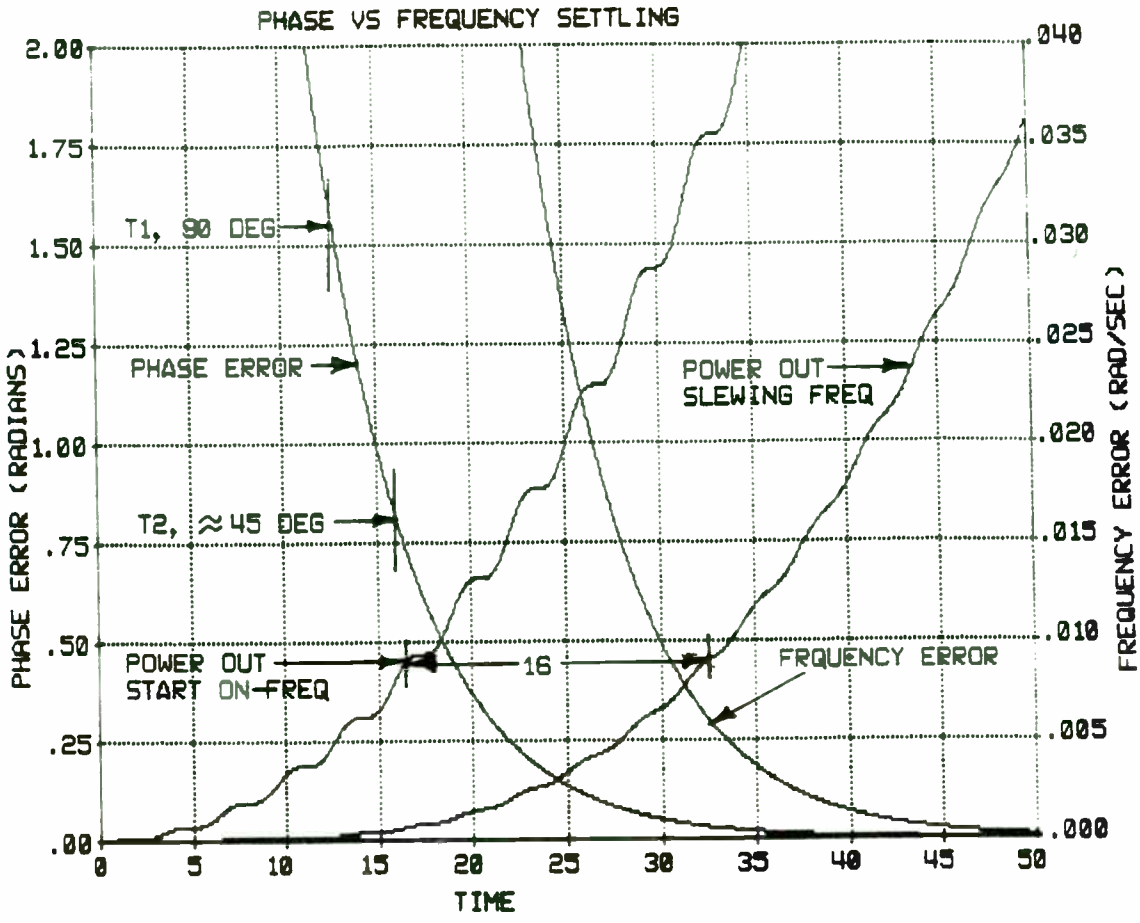


Fig. 2 - Simulation Results

The plot in figure 2 is from 0 to 50 time units (seconds in this case). The rising line marked "power out, start on-freq" is the filter output power build-up from a sudden application of the correct frequency and stable phase, as the high-Q filter integrates the input signal. This is the filter response with an input of zero settling time.

The next rising line marked "power out, slewing freq" is delayed because of the input settling time. The delay is almost uniform over time, as these output levels illustrate:

Times(sec)		Delay(sec)
12.5	27.5	15
17.5	33	16.5
25	42	17

This indicates that the total hopping period has lost about 16 time units of performance under these settling conditions. Now let us compare this delay with the frequency and phase settling characteristics.

The phase error has settled to within 90 degrees at t_1 = 13 sec. At t_2 = 16 sec. the phase error is about 45 degrees, which corresponds to a system performance loss of 16 seconds delay of signal reception. Note that the time t_2 is 3 sec. later than t_1. This will be checked in the analysis section.

The frequency error is shown according to the right-hand scale. Starting at an error of 4 rad/sec, it finally reaches the passband of .01 rad/sec in *30 sec.* or almost twice as long as the beginning of signal reception!

Comments on the Simulation Results

This simulation illustrates the following very important points:

(a) By far the largest contributor to performance loss is the time (t_1) spent *before* the incoming phase comes to within 90 degrees from its final settling value.

(b) A small additional time (from t_1 to t_2) is spent just after the phase reaches its final 90 degrees, until the equivalent time of full system performance.

(c) The freqency error ω_e is the loop bandwidth times the phase error θ_e: Note that faster tuning (wider

bandwidth) loops have higher frequency errors for the same phase error. The basic first-order equations are:

$$\text{eq(1)}\ \theta_{e(t)} = \frac{\Delta\omega}{\omega_n} e^{-\omega_n t},\ \text{eq(2)}\ \omega_{e(t)} = \Delta\omega e^{-\omega_n t}$$

where $\Delta\omega$ = freqency step and ω_n = pll bandwidth.

Higher order loops have more complex expressions but the final phase settling is similar. At time t2, the instantaneous frequency error is the loop bandwidth times the phase error, or 0.2 * 0.785 = 0.157 rad/sec. This is over 15 times the filter upper band edge of 0.01 (above the band center of 1)! This is a subtle point because it is somewhat of a brain-teaser that even though the incoming instantaneous frequency may be much different from the steady-state-bandwidth of the receiving system, the filter does in fact derive useful energy from the incoming signal!

(d) One explanation for this paradox is that the transient (as opposed to steady state) frequency response of filters increases towards infinity as the applied transient time decreases from 1/bandwith to zero.

(e) Another way of looking at this is that the faster a frequency changes, the wider its spectrum becomes due to the FM sidebands. The sidebands are already driving the filter even though the "carrier" ("instantaneous frequency") is not "there" yet.

(f) Still another intuition is that the filter is driven by phase. During the earlier part of settling the phase is rotating completely around the circle (because the frequency difference produces a beat note) and so the net effect is zero. The filter output starts rising when the periods of the incoming signal are approximately equal so that they continue to be in phase with the previously established filter ringing. The instantaneous frequency is the derivative of phase, so if the phase is settling rapidly the frequency is still outside the passband! In this sense a filter acts like a phase detector.

Settling Specifications

It is desirable, in terms of overall loss, for a frequency hopping system to have a phase characteristic such that the incoming and reference phases come to within 90 degrees as fast as possible. The time spent later on in matching the two frequencies is not nearly as important for minimizing total system loss.

A frequency hopping system should be specified in terms of overall performance loss given a hopping rate and related parameters. The commonly used specification which requires the receiver to be within a certain frequency offset of the desired frequency is not a good one, because it can be met with many different combinations of design parameters each of which combinations results in a different performance loss for the overall system!

Simultaneous phase and frequency settling specifications are not advisable because they effectively specify the phase-locked loop bandwidth, which may very likely interfere with obtaining the best tradeoffs between tuning speed, noise, and spurious levels, as well as cost, size, power, etc.

Measurement of "Instantaneous" Frequency

The concept of instantaneous frequency is hazy and should be approached with caution. It seems wise to reserve the word frequency for steady state conditions, unless very special precautions are taken. Measurement of the instantaneous frequency in a fast hopping system can be very difficult, which can lead to traps for the unwary. For example, if the synthesizer "slewing" frequency is gated at very short intervals to a narrow band filter, such as a spectrum analyzer with a 10 Hz bandwidth, it can appear to be within the 10 Hz band when the filter is simply ringing because of the sudden gating impulse, even when 100 kHz away!

Instantaneous frequency has also been measured by the slope of the phase error. A synthesizer phase error is easily measured by mixing with the desired reference frequency and observing the beat note on an oscilloscope. The reference phase is adjusted until the phase error crosses the zero center line at the specified settling time. From a measured small $\Delta\phi$ ($<<1$ radian) and corresponding Δt, the frequency = $\Delta\phi/\Delta t$. For small frequency errors this can be a very difficult measurement, even buried in the noise, which led to the suspicion that it may not have been a good system parameter.

Analysis

Various methods of analysis of system performance loss have been used, all of which have given approximately the same results. An analysis by Caloyannides [ref. 1] used integration of (sin x) / x terms which is difficult mathematically.

An analysis by Schoenike [ref 2] used a simpler method which gives approximately the same results. The model is an integrated output of an orthagonal synchronous detector, whose equation for the output voltage is eq(3):

$$E_o = K\sqrt{\left[\left(\int_0^T \cos\theta_1(t)\cos\theta_2(t)dt\right)^2 + \left(\int_0^T \cos\theta_1(t)\sin\theta_2(t)dt\right)^2\right]}$$

Trigonometric substitution results in sum and difference terms. The sum terms can be eliminated because it is assumed that there are many phase oscillations over the period T, so a very close approximation is eq(4):

$$E_o = \frac{K}{2}\sqrt{\left[\left(\int_0^T \cos[\theta_1(t)-\theta_2(t)]dt\right)^2 + \left(\int_0^T \sin[\theta_1(t)-\theta_2(t)]dt\right)^2\right]}$$

By a proper choice of $\theta_2(t)$, the sine integral can be made zero. Then

eq(5) $$E_o = K_1 \int_0^T \cos[\theta_1(t)-\theta_2(t)]\,dt$$

To evaluate this integral, θ_2 is set equal to the final angle at which θ_1 eventually stabilizes. Also, the period between samples of θ_1 must not exceed π, so the numerical integration should not begin until the phase difference between adjuacent samples is less than π . This effectively sets the integral to zero up to the first usable sample, because the positive and negative half-cycles up to this point tend to cancel each other.

A phase-locked loop phase settling characteristic (as previously defined) is substituted for the phase difference in eq. 5. The integration result is divided by the total time T to find the actual voltage to ideal voltage ratio. This loss ratio is now:

eq(6) $$Loss = \frac{t_{\pi/2}}{T} + \int_{t_{\pi/2}}^{T} \cos\left(\frac{\Delta\omega}{\omega_n}\, e^{-\omega_n t}\right)^2 dt$$

eq(7) $$Loss = \frac{t_{\pi/2}}{T} - \frac{\Delta\omega^2}{2\omega_n^2}\int_{t_{\pi/2}}^{T}\left(e^{-2\omega_n t}\right)dt$$

eq(8) $$Loss = \frac{t_{\pi/2}}{T} + \frac{1}{4\omega_n T}\left(\frac{\Delta\omega}{\omega_n}\, e^{-\omega_n t_{\pi/2}}\right)^2$$

eq(9) $$Loss = \frac{t_{\pi/2}}{T} + \frac{\pi^2}{16\omega_n T} \approx \frac{t_{\pi/2}}{T} + \frac{0.6}{\omega_n T}$$

The loss is in two parts:

(1) the time from the start of frequency switching (t = 0) until the final 90 degrees from the final phase, and

(2) an additional time related to the loop bandwidth. This is commonly a much smaller time, such that the equivalent phase for a "settled" condition may approximate 1 radian.

The previous simulation had no specified T, so the actual settling times will be found by multiplying eq. 9 by T, resulting in the time t_1 plus the time interval $t_2 - t_1$, or t_{add}, the additional settling time after the 90 degree point:

eq(10) $$t_{add.} = \frac{0.6}{\omega_n}\,\text{sec.} = \frac{0.6}{0.2} = 3\ \text{sec.}$$

which agrees with the simulation.

Conclusion

The frequency synthesizer phase settling characteristics were found to be a critical factor in the performance of frequency hopping systems with FSK type of modulation with integration over each hop period. When the sysnthesizer settles in an exponential manner (such as with PLLs), the effective system performance starts after two time intervals have elapsed. The first is from the start of frequency change to the last time that the phase error (relative to final settled phase) is 90 degrees. An additional small time interval is 0.6τ (1 / PLL bw). Typically this results in a settling time until the phase error is approximately 45 degrees.

Other conditions can also affect the system performance, but frequency error has no direct effect. In some cases specifiying frequency settling may lower system performance or compromise the synthesizer design, because of the many parameter trade-offs in PLL design.

References

1 Caloyannides, Michael. A., *An Analytic Evaluation of Synthesizer-Switching Performance Loss*, Rockwell International Working Paper, 1981.

2 Schoenike, Edgar. O., *Synchronous Detection of a Slewing Signal Frequency*, unpublished paper, 1976.

Linear Frequency Modulation — Theory and Practice

Bar-Giora Goldberg, MSEE
Henry Eisenson
SCITEQ Electronics, Inc.

Introduction

Linear frequency modulation (FM), or a "chirp," is a swept frequency change over some bandwidth. An ideal chirp has a perfectly constant rate of change (time vs. frequency), and is generated by a means that makes it perfectly reproducible. Stimulated by emerging radar requirements to distribute energy across a range of frequencies, early LFM signals were generated with analog devices such as a voltage controlled oscillator (VCO) or (later) by surface acoustic wave (SAW) resonator.

This paper will address chirp signals in general, and present one aggressive all-digital mechanism for generating them with a combination of linearity and operating bandwidth not previously achieved.

History

In the late 40's and early 50's, as pulse radar technology grew out of its infancy, a problem became visible. Radar range depends on $\frac{E}{N_0}$, where E is the pulse energy , $E = P \cdot T$, (P is the received power proportional to the transmit power, T is the pulse duration) and N_0 is the receiver noise density, and resolution depends on the signal bandwidth. These two values are related diagonally, as improving E makes it necessary to increase T and an increase in T increases the pulse bandwidth and therefore resolution.. One solution was to modulate the pulse, making pulse bandwidth dependent upon the modulating waveform rather than pulse length of (approximately) $\frac{1}{T}$. Linear FM (LFM) was developed and adopted long ago, and it remains useful today in many radar applications. Relatively recently, the utility of LFM in non-radar applications has been recognized and explored, and. chirp is now important in communication, semiconductor process control, sonar, seekers, range measurement, simulation, test equipment and other system categories. VCO solutions have analog accuracy, and problems with linearity and repeatability. Those effects are lessened with SAW devices, but SAW is neither programmable now deterministic.

In 1989, Sciteq began work on an all-digital technique to generate LFM, using direct-digital synthesis (DDS) technology. A DDS consists of an accumulator stage to correlate the clock with a control word to produce a phase word, a memory to map phase information to amplitude, and a digital-to-analog converter (DAC) to produce the analog waveform. The problem was that architectures selected for typical accumulators included a great deal of latency partly due to pipelining (to conserve power), which meant that a new frequency could not be loaded until the previous one had propogated through the logic. The resulting delay was acceptable for audio/sonar frequencies, but not at microwave. In addition, some means had to be developed to change the frequency control word to the accumulator at a rate commensurate with the desired output ramp. These deficiencies mandated development of new DDS integrated circuitry that would provide a cost-effective approach to an idealized chirp signal.

The goal therefore was to develop a fully digital, programmable, chirp generator to enable the digital of F_{start}, slope, and phase modulation — all fully synthesized. The solution must provide total parameter control, with digitally deterministic repeatability.

By 1990, necessary digital and data conversion components had already been developed, or were being developed, either by Sciteq or other agencies (Sandia National Laboratories, for one), and for the first time it appeared possible to produce the industry's first wideband digital chirp synthesizer. Accordingly, Army Research Laboratory (then Harry Diamond Labs, in Adelphi, MD), let a Small Business Innovative Research (SBIR) Phase I contract to Sciteq to determine the practicality of meeting theoretical goals for the Army's next-generation battlefield surveillance radars. Specifically, the ARL group working on Synthetic Aperture Radar believed that a digital signal generation solution was important to future program objectives. Sciteq's interface for this program, and sponsor, was and remains Barry Scheiner, of ARL.

Phase I defined a practical path to the desired goal, leading to a Phase II contract award. Most SBIR projects are considered high risk, but this development was successful and prototypes have been delivered to the Army and to other agencies interested in evaluating ideal linear FM signals.

The nature of the SBIR program permits the contractor to retain control over the new technology, and to productize the results where possible, hence the availability of Sciteq's direct-digital chirp synthesizer (DDCS) to the industry. That product, designated the DCP-1, is a DDCS that produces linear FM signals over a band of more than 230 MHz, with linearity approaching an ideal level previously defined only by mathematics and never actually observed.

Waveform Properties

A substantial amount of theoretical work was done to calculate the properties of LFM signals, i.e., power spectrum, auto-correlation, etc., and the characteristics of a theoretically perfect LFM signal were defined. A general description of such a signal is given by:

$$s(t) = A \cdot \sin(\alpha t + \frac{\beta}{2} t^2 + \varphi_0)$$

If the signal bandwidth is designated as W, then the product, $W \cdot T = TB$, becomes the time bandwidth product in radar applications or processing gain in spread spectrum communication. For $W \cdot T = TB >> 1$,, the power spectrum is flat and the signal energy is distributed almost equally across the bandwidth, W. Signal auto-correlation is given by:

$$R(\tau) = \left(\frac{\sin X}{X}\right)^2 \text{ where } X = \pi \cdot \tau \cdot W = \pi \cdot \tau \cdot \frac{TB}{T}$$

It can be shown that the signal auto-correlation has relatively high sidelobes with the worst case of -13.5 dB (approximately $20 \cdot \log\left(\frac{\sin X}{X}\right)$ for $X = 1.5 \cdot \pi$). This parameter can be improved by adding the complexity of either amplitude modulation or phase modulation (pulse weighting).

Applications

LFM is used for fuzing, altimetry, ranging, pulse Doppler radars, synthetic aperture imaging radars and more. The better the linearity of the LFM signal the better the performance of the system, but even the chirps generated by analog circuitry were better than non-chirped operation. With the advent of DDCS technology, with its inherent deterministic linearity, imaging resolution of SAR and performance of other chirp-dependent systems were substantially improved. For those same reasons, LFM provides advantages to range measurement systems, for altimeters, flight control and similar applications.

Compressive receivers are used to scan the spectrum, and unlike typical spectrum analyzers, which sweep the spectrum and find signals that are within the analyzer bandwidth during the time of the sweep only, these receivers see ALL signals that occur within the sweep time. In this respect, a compressive receiver operates like a real time DFT analyzer. An ideal compressive receiver depends upon a linear and fast chirp over the desired band of reception.

In semiconductor production, wafer perfection is a critical issue that influences yield, and therefore has serious economic impact. Surface anomalies and contamination are usually detected using a laser beam, which is reflected off the wafer surface and scatter characteristics evaluated to determine surface characteristics, resulting in a map of the wafer showing imperfections, contamination, and therefore usable dice. In most such applications, the wafer is moved mechanically in one dimension, and the laser is modulated by a Bragg cell driven by a linear FM signal (produced from an expensively-compensated VCO). The linearity of the FM signal is one of the factors that determines resolution of the system along the appropriate axis, and also affects the "guard" area around imperfections that are detected (because of inherent error). The more accurate the signal (the more controllable the laser position) the smaller the guard area and the higher the yield.

Digital LFM Synthesis

Evolution of digital technology has allowed certain Direct Digital Synthesizer (DDS) implementations to operate at sufficient speed to produce output bandwidths sufficient for the above described applications.

The use of DDS disciplines brings with it cardinal improvements in the waveform features such as:

- The signal is synthesized and therefore every pulse is identical to the previous one.
- The chirp linearity approaches the limits of measurement.
- Phase manipulation is digitally accurate and is available at almost no additional cost.
- Control of parameters such as start frequency, stop frequency, chirp rate, on/off, etc. are completely and inherently deterministic and accurate.

The output of the digital process is represented as:

$$s(n \cdot t_o) = \sin\left(\frac{\beta\,(n \cdot t_o)^2}{2} + \alpha(n \cdot t_o) + \varphi_o\right)$$

The general structure is that of a dual accumulator. Since an accumulator is a discrete integrator and the requirement is to generate a quadratic functinon, two accumulators are used. The output of the first accumulator is the instantaneous frequency and the frequency adder allows the setting of a start frequency. The F output allows the monitoring of the instantaneous frequency. The input of the first accumulator is therefore β in the equation and the frequency input is α. The output of the second accumulator is the signal's phase and therefore can be phase modulated by another adder. This input is equivalent to the term, φ_0, in the equation.

Such structure can be implemented in CMOS at low clock frequencies, and in high speed ECL and GaAs technologies up to 1000 MHz, though the nature of a DDS limits output to less than half the clock.

In the existing design, the chirp chip and phase adder (implemented in one device) are followed by a SINE ROM, a DAC (digital to analog converter), and a low pass filter. Because of the nature of the output spectrum from a DDS $\left(\frac{\sin x}{x}\right)$ and the group delay of the filter, amplitude and group delay equalization is included to improve the result.

Practical Implementation: the DCP-1

The Sciteq model DCP-1 is a direct-digital chirp synthesizer (DDCS) that is clocked at 500 MHz and therefore generates output frequencies from DC to 230 MHz (limited by Nyquist and the low pass filter characteristics). The basic chirp generation function is achieved by three devices, a double-accumulator, a memory, and a digital-to-analog converter.

In the double accumulator, both accumulators are 24 bits in size, thus yielding a minimum step size of ~29.8 Hz at a clock rate of 500 MHz. The frequency and phase accumulator functions are integrated into one device, developed by a Sandia National Laboratories program under the leadership of Bruce Walker. The part includes not only the specified accumulation functions, but also 12-bits of phase control and a time-equalized 8-bit frequency output that supports system timing. The memory device uses a patent-pending algorithm (Sciteq's) to map the phase output of the accumulator to digitally-defined amplitudes, also at a 500 MHz rate or better. The digital output of the memory is considered near-perfect, with digital error supporting a spurious response better than 70 dB below the carrier, so it is the digital-to-analog converter (DAC) that limits the spectral purity of the system. Initial DCP-1s used a 12-bit GaAs part developed by a consortium including GE, Sciteq, Motorola, and Hughes, and typical spurs are at the -45 dBc level.

In the LFM mode, the DCP-1 updates frequency at a 500 MHz rate, which means that a new frequency is synthesized every two nanoseconds. The slowest chirp rate is ~15 kHz/μsec (~30 Hz times 500, since there are 500 steps in each microsecond). For a full band sweep, this would take approximately 13.4 msec (230,000,000 ÷ 29.8 • 2 nsec = 15.4 msec). Faster sweeps are possible, limited by acceptable resolution (as an extreme, at 230 MHz resolution it's one full chirp – a single step in two nanoseconds). A chirp rate of 10 MHz/μsec is practical if it is desired to cover a 50 MHz bandwidth in 5 μsec (5 μsec ÷ 2 nsec for the number of steps gives a required resolution of 20 kHz).

The synthesizer is controlled by loading two registers — start frequency and chirp rate (or step size). When the chirp begins, the step size will be added to the start frequency every 2 nsec. At any point during the chirp it is possible to change the chirp rate to produce different time:frequency relationships. A negative value in the chirp rate register (2's complement) will produce a sweep starting at a higher frequency and moving lower, thus supporting complete manipulation of all output parameters.

In addition, 12 bits of phase control permit compensation of the response during the sweep to reduce side lobes. This may be updated at a rate limited only by speed limitations of TTL logic. Phase control adds another dimension of flexibility by permitting the control of phase from pulse to pulse, which permits accurate matching of signals.

Applications and Experimental Results to Date

The DCP-1 is now used by a variety of systems, including (upconverted/multiplied) two millimeter-wave seeker programs, four synthetic aperture radar systems, two electronic warfare programs, and (in baseband) at least one wafer process control system.

So far, system developers have reported favorable results. Spurious signal level was initially a concern (as is the case in most DDS applications), but one unpredicted result of experimentation is that discrete spurious signals seem to be integrated into the general output, and have little result on overall system performance.

Linearity is within quantization levels, therefore for broadband chirps the errors are smaller than conventional measurement techniques can detect. Initially, repeatability was evaluated by delaying the output of the DCP-1 and then comparing that signal with the original; the result was measured on an HP 3561A FFT and found to be virtually perfect. Linearity testing was conducted using the Racal-Dana 2351 Time Interval Analyzer, and the HP 5373A Modulation Domain Analyzer.

The results include both time vs. frequency data and a histogram displaying frequency distribution, and again support the contention that the linearity is perfect within the limit of granularity.

SUMMARY

Several niches in the RF industry can benefit from linear FM signals. Such waveforms have been generated using various analog means, with results that improve system performance despite the ambiguities of the analog design, and engineers in these niches seek to generate signals with improved linearity and predictability. Sciteq, sponsored by the U. S. Army and in cooperation with Sandia National Laboratory, has developed a linear FM generator that synthesizes a new frequency each two nanoseconds. Based upon direct-digital synthesis, the new technology creates opportunities for optimization of synthetic aperture radar, altimetry, ranging, seeker, and even process control systems.

RF ACTIVE DEVICE MODELING FOR CAD, A COMING NECESSITY

Gary Roberts
Hewlett Packard Company
Santa Rosa Systems Division
1400 Fountaingrove Parkway, Santa Rosa Ca, 95403

INTRODUCTION

As the world electronics industry has reoriented to focus on commercial and consumer electronics, products in the RF frequency range have become extremely important. Further, the competitiveness in these markets demands that reduced cost and time to market become primary goals in product development.

Accomplishing these goals requires the ability to consistently simulate the performance of circuits before they are built. Key to achieving a single design cycle, and the associated cost/time reduction, is the availability of accurate, up-to-date device models.

In the past, the importance of accurate models in achieving a single design cycle has been recognized. However, actual statistics illustrating the importance are rare. Figure 1 shows data from a world-class foundry for 36 MMIC product designs completed over a ten year period. Each design experienced a number of iterations, and the reasons for each iteration have been analyzed and plotted. The lack of adequate models was the largest contributor to added iterations, accounting for over 30% of the total. The breakout of data for the last two years indicates it is still the major cause (40%).

MODELS AS A COMMON LANGUAGE

Achieving a single design cycle requires the total engineering efforts of the device designer, the fabrication engineer, and the circuit designer, each working individually and in combination. In today's environment, device models have come to serve as indispensable tools to aid the success of each engineer as well as to provide a common language that facilitates the summing of their efforts. For example, the circuit designer now has a broad range of CAD tools available, yet it is accurate device models which allow these tools to be used to their maximum extent. Circuit performance may now be optimized with select model parameters, and specification of circuit needs is communicated with model parameter values, thereby providing targets for device/technology development. These same model parameters then provide both targets and quantitative measures for the fab engineer to monitor and control the key process steps that assure product consistency.

WHAT IS A MODEL?

To achieve its ideal, a device model should provide a representation of the electrical performance of the device which can then be used either for circuit simulation or as a detailed measure of transistor performance. However, there is not yet one "universal" transistor model type which ideally suits all purposes. Accordingly today, models are divided into three broad categories: physical models, empirical models, and data-based models (Figure 2). Each has its advantages and disadvantages.

Physical models derive device performance from a two-dimensional analysis of the electron dynamics in the device structure (ref1). Electrical performance is obtained from a detailed solution of Poisson's equation while evoking current continuity, and energy and momentum conservation. Performance is provided in terms of device geometry, doping concentrations, and material constants. This class of models is particularly useful in directly relating device physics, geometry, and fabrication details to device performance. Thus it provides the best insight into how to design and fabricate a better device.

Physical models, however, often must be simplified to provide reasonable circuit simulation times. In doing so, they lose varying amounts of accuracy. Recent work to overcome this deficiency for GaAs FET and HEMT devices has shown promise(ref1). However, their usefulness to circuit designers is still limited because they still require knowledge of fabrication details which are not always available, or are considered to be proprietary by component vendors.

Empirical models are the most widely used today for circuit simulation. They represent the best compromise for use by both the device and circuit designers. Examples of this class of models are the industry standard Gummel-Poon model for BJT devices (ref2), and the SPICE MOS models for MOSII and III(ref3) and BSIM (ref4). Here the device is represented by an equivalent circuit schematic whose form and components are derived from both device physics and

empirical equations. Device nonlinear behavior is modeled with parameterized empirical equations that are functions of the various control voltages. Parameter values are "extracted" by curve-fitting simulated results to measured data. Model accuracy depends on the accuracy of the measured data, the fitting process, and the complexity of the empirical functions. Often, the equations that describe a particular schematic component are intentionally made simple to facilitate model use in circuit simulators. This simplification bounds model accuracy and limits the voltage or current ranges and the circuit applications for which a particular model may accurately apply.

Data-based models are the newest form of nonlinear device model. They have been created specifically to address the need for higher accuracy in circuit simulation. For this class of models, the measured data itself is used to generate nonlinear model functions. HP's Root Models for FETs (ref5), diodes (ref6) and MOS devices are examples. Because the model is explicitly constructed from the measured data (S-parameters vs bias) it is technology and process independent. This class of model provides the highest level of overall accuracy for use in circuit design, but is not useful in providing insight into how to design and manufacture higher performance devices.

HOW IS A MODEL OBTAINED?

Because empirical models represent a good compromise for use in both circuit design and device development/fabrication, it is informative to review the key steps involved in obtaining a parameter set for a typical SPICE model. For our example let's use the Gummel-Poon (GP) model for a silicon BJT (ref2)

The schematic equivalent circuit derived by Gummel and Poon for an NPN BJT is shown in Figure 3a. It was obtained from a consideration of the device physics and geometry. The model DC node currents, IB and IC, were obtained by summing the contributions of current flowing through each of the diode and current sources (Figure 3b). Values for the empirical model parameters (such as the diode ideality factors nF and nR) are empirically determined in the modeling process by adjusting their values until the simulated currents agree with the measured currents over specified ranges of voltage. Through the use of select DC and RF measurement sequences and techniques, and appropriate test hardware and test and modeling software, complete BJT SPICE parameter sets (Figure 4) can now be obtained in as little as two hours (ref7).

Once the SPICE model parameter set is obtained, device S-parameters can then be simulated over a wide range of bias conditions from near cutoff, through the active region, to near saturation (Figure 5). The parameter set may then be set up as a device model file in a model library, ready for use by the circuit simulation software. Thus, through the use of fast efficient modeling methods and systems, it is now possible to build and maintain model libraries in support of CAD circuit simulation tools.

WHERE, THEN, DO MODELS COME FROM?

For the circuit designer today, device modeling information to support CAD tools is available from three major sources: model libraries, a model extraction service, or a modeling group/system set up in support of the circuit design. Since information from the latter two sources ultimately ends up in some kind of custom library made by combining custom and commercial model data, the information in such a library may take many forms and vintages.

The information found in commercial libraries provided by CAD vendors is in two tiers. The first tier consists of S-parameters data at a fixed bias, the same data typically available from data sheets or data books. CAE software companies have simply repackaged it into a form more directly usable within their simulation software. The data is available on a broad range of devices, but its usability is limited to small-signal applications at the same bias at which the initial data was taken. The second tier is represented by complete sets of extracted model data. This data exists for many fewer devices, but the data is much more usable in that it can be used to simulate device operation for a wide range of nonlinear circuit applications over a wide range of bias conditions. It thereby provides the circuit designer more flexibility in optimizing circuit performance.

Recently, CAD vendors have also begun to include device layout data (when available) in addition to the electrical data. This facilitates circuit layout as well as the simulation of circuit electrical performance.

Model data contributed from commercial libraries has both pros and cons. On the one hand, it is relatively inexpensive, while on the other it is most often encrypted in the CAD vendor's code and not available for inspection. In this form its accuracy or age is difficult to assess. Commercial libraries usually contain only the most popular commercial devices and do not include custom devices or models sometimes desired and used by designers. Even for standard devices, the designer may not have control over how the data was generated (e.g. bias points), or what device model may have been used to fit the data.

A parameter extraction service is an excellent alternative to supplement the information contained in a commercial library. Here the designer may pay the service to have a specified device (or devices) modeled by an experienced modeler. The fees charged are typically economical for a small number of devices; the designer may often specify a preferred model type; and the service provides a means to obtain models for custom or newly-developed devices that have not yet found their way into existing libraries. These

services should not, however, be viewed as a primary method to obtain a custom library. When a large number of devices are to be modeled, the service may become prohibitively expensive. Also, long lead times may be experienced depending on the backlog at the service. These delays are additive to the circuit design cycle, and increase the time to market.

As the quantity of devices modeled becomes large, or as competition grows, there is a point at which companies are finding it economically or competitively advantageous to establish their own in-house modeling capabilities. These functional areas may be set up in support of a variety of needs: those of in-house circuit designers, those of component vendors who supply model data along with their components, and those of in-house device design and manufacturing. With this alternative, the designer may exercise more control over the modeling process, can obtain models on a timely and periodic basis, and can quickly obtain models for new or custom devices. More companies are realizing that ready access to and use of this model information represents a competitive weapon that can be used to differentiate them from the competition.

While there is a larger initial cost to the establishment of such an internal modeling capability, represented by the purchase of the hardware and modeling software and the establishment of the engineering expertise, even more rewards are to be reaped in the potential reduction of design cycles and time to market.

MODELS (AND LIBRARIES) ARE TIME-PERISHABLE

In today's dynamic business and technical climate, the advantages gained from a custom device library are not long-lived unless the information is continuously reviewed and updated. Both standard and custom devices must be characterized, models generated, and information archived. However, this alone does not assure that the model data, once created will continue to be useful. For data to remain useful, it must represent what can currently be manufactured. But because device technology is continually evolving, and fabrication processes are time-varying, there is a continued requirement for new models to be developed and model data to be periodically reverified. Models and data that reside in libraries must be viewed as time perishable.

To illustrate this point, Figure 6 plots the normalized value of the gate capacitance, Cgs, for a GaAs MESFET device extracted using a simple FET model. The data represents measurements from over 1300 FETs from 56 different wafers that were fabricated using the same foundry process over a period of eleven months. The distribution of Cgs values from FETs on each wafer are represented by each box and its outlying points. The slow time variations of Cgs are due to slow changes in the overall fab process. The abrupt changes were due to an intentional change in one portion of the process to retarget the capacitance. The Cgs data illustrates that, once created, device model data needs to be reverified with a frequency that relates to the dynamics of the technology and/or the fabrication process stability.

LET THE DESIGNER BEWARE

Library model data should, therefore, be used with caution. The designer would be wise to question its age and accuracy before use. Not all the data that exists in CAD libraries, or that is available from component vendors' data sheets, may be adequate to support the full capabilities of CAD software tools. As we have seen before, some of the data may be in a limited form, (DC and small-signal S-parameters only). Also, some may be out-of-date (no longer representing what is manufactured), or some may be inaccurate (having been taken with less-controlled methods in years past). The use of this data may provide "first pass" circuit designs, but may also necessitate further design cycles to adjust the circuit performance to acceptable levels.

Accordingly, designers are increasing requests to component vendors (both captive and commercial) to supply more accurate, up-to-date model data, and to do so on a more regular and ongoing basis. In defense of these same vendors, many are now establishing or upgrading modeling capabilities using systems like that shown in Figure 7. These systems typically consist of a suite of hardware and test and modeling software. The systems interface to the device in a variety of ways, with probes for die in wafer form, and fixtures for die in chip or packaged form. Older versions of these systems have been put together using hardware from a variety of commercial sources, and modeling software which is either commercial or (in most cases) "home grown".

For those that are upgrading or establishing new modeling facilities, complete modeling systems like the HP system in Figure 8 are now available. These systems provide a variety of test hardware combinations, and also include test/modeling software, documented modeling techniques (instrument cal, device biasing, probing and fixturing), and total system support for both the hardware and the modeling software.

CONCLUSIONS

Component vendors are realizing the benefits of these systems in the form of added component sales because designers prefer parts that are supplied with accurate, updated model data. Indeed, these same vendors now view the supplying of this data as a competitive weapon. In the future, it will become an absolute necessity as more of their competitors are able to supply reliable component models.

While this greatly benefits the circuit designer, not all sources of desired components will be able to supply data in as timely a manner as may be required to meet future time-to-market goals. For this reason, design groups are still choosing to establish their own modeling capabilities in direct support of their design efforts, using these capabilities to create their own custom libraries of component models. The benefits are added performance, reduced product cost, and reduced time-to-market, all competitive weapons in an increasingly competitive electronics market place.

REFERENCES:

1) Snowden, C.M., MESFET Modeling, Compound Semiconductor Device Modeling, Springer-Verlag, London, 1993, pp 26-55.

2) Gummel, H.K., and Poon, H.C., "An Integrated Charge Control Model of Bipolar Transistors", Bell System Technical Journal, 49, 1970, pp 827-852.

3) Vladimirescu, A., and Liu, S., "The Simulation of MOS Integrated Circuits Using SPICE2", Electronics Research Laboratory Rept. No. M80/7, University of California, Berkeley, 1980.

4) Sheu, B.J., Scharfetter, D.L., and Poon, H.C., "Compact Short Channel IGFET Model (CSIM), Electronics Research Laboratory Rept. No. M84/20, University of California, Berkeley, 1984.

5) Root, D.E., "A Measurement-Based FET Model Improves CAE Accuracy", Microwave Journal, September 1991, pp 126-136.

6) Root, D.E, Pirola, M., Fan, S., and Cognata, A., "A Measurement-Based Large Signal Diode Model, Automated Data Acquisition System, and Verification with On-Wafer Power and Harmonic Measurements", IEEE MTT-S International Microwave Symposium Technical Digest, 1993, pp 261-264.

7) Dunn, M., "Device Model and Techniques for Fast, Efficient Nonlinear Parameter Extraction of BJT Devices", Asia-Pacific Microwave Conference Technical Digest, 1993 (to be published).

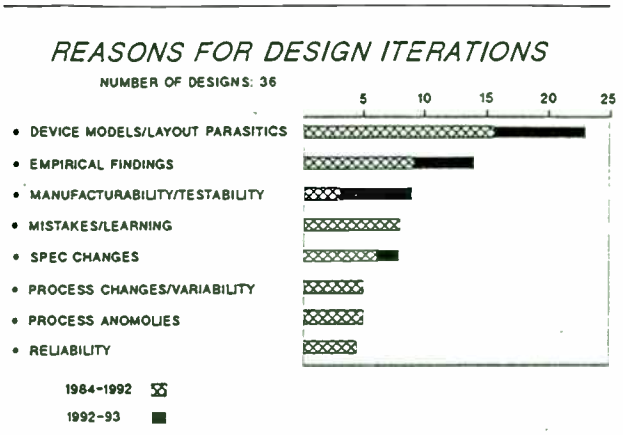


Figure 1. Design iterations

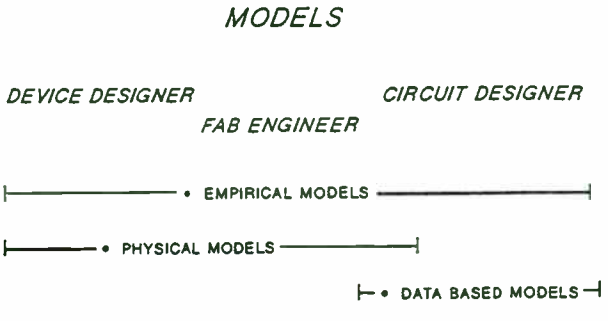


Figure 2. Model types

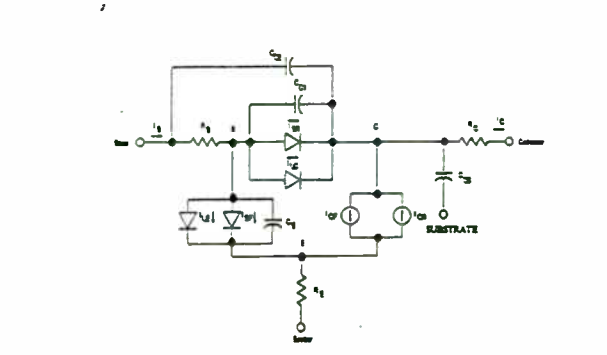


Figure 3a. BJT model schematic

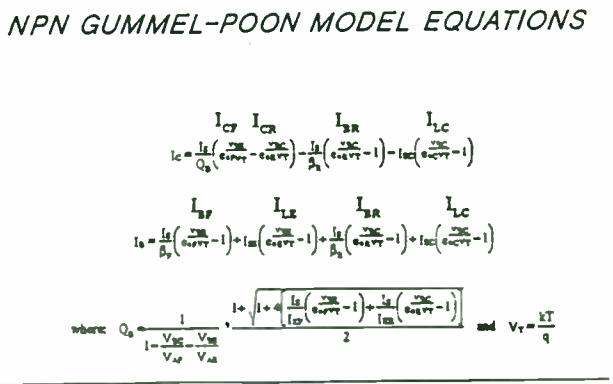


Figure 3b. BJT model equations

GUMMEL-POON PARAMETERS

SYMBOL	DESCRIPTION	SYMBOL	DESCRIPTION
IS	Transport saturation current	RBM	Minimum base resistance
NF	Forward ideality factor	RE	Emitter resistance
VAF	Forward Early voltage	RC	Collector resistance
IKF	Forward knee current	CJE	Base-emitter zero-bias capacitance
BF	Forward beta	VJE	Base-emitter built-in potential
ISE	Base-emitter saturation current	MJE	Base-emitter grading coefficient
NE	Base-emitter ideality	CJL	Base-collector zero-bias capacitance
BR	Revrse beta	VJL	Base-collector built-in potential
NR	Reverse ideality factor	MJL	Base-collector grading coefficient
VAR	Reverse Early voltage	TF	Forward transit time
IKR	Reverse knee current	XTF	Coefficient of TF bias dependence
ISC	Base-collector saturation current	ITF	Models TF dependence on IC
NC	Base-collector ideality factor	VTF	Models TF dependence on VBC
RB	Zero-bias base resistance	PTF	Excess phase of TF
IRB	Current where base resistance falls halfway to its minimum value	XCJC	Models distributed nature of base
		FC	Models transistion from junction of diffusion capacitance

Figure 4. Gummel-Poon BJT parameters

MODELING RESULTS

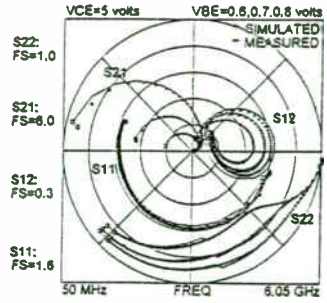


Figure 5. Verified S-parameter data at three Vbe values

EXTRACTED FET GATE CAPACITANCE
Cgs (normalized) vs TIME

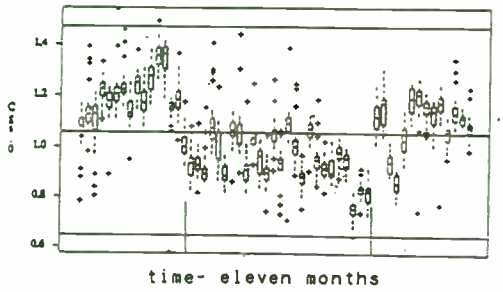


Figure 6. Modeled capacitance vs time

RF MODELING SYSTEM

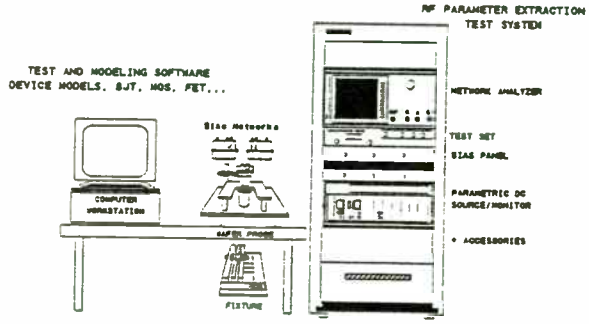
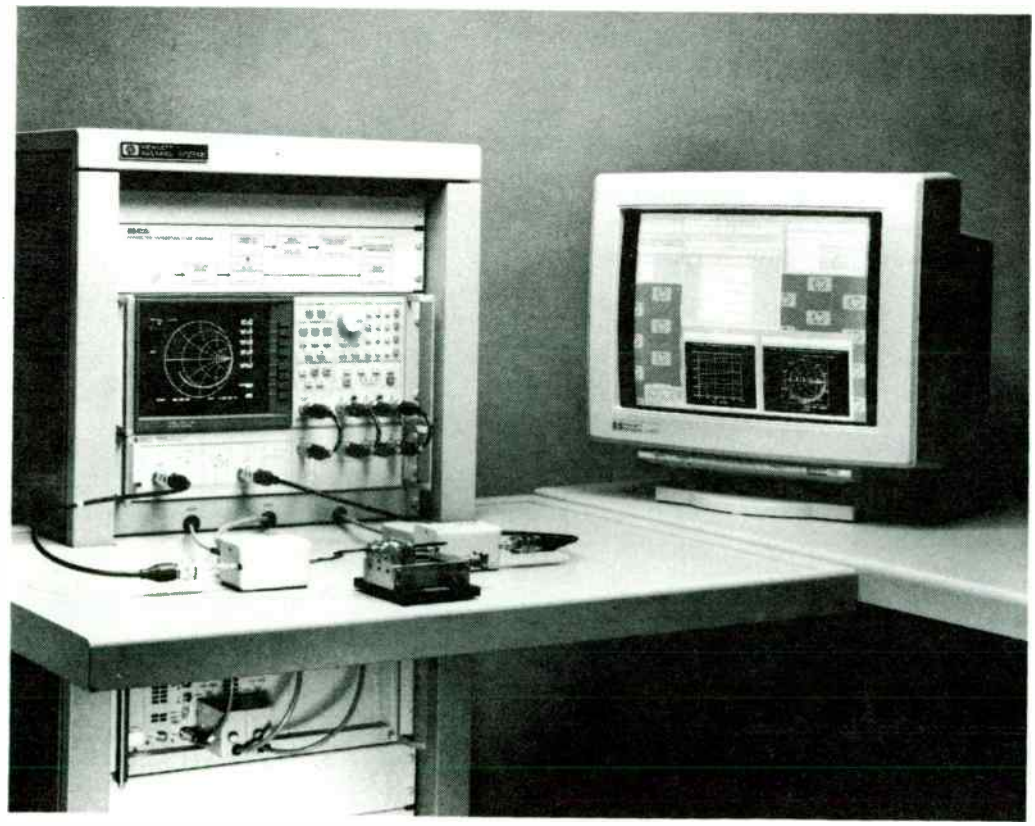


Figure 7. Typical modeling system configuration



Regression-Based Algorithms for Inductor Modeling

Edmund (Joe) Tillo II, Ph.D.
Ford Motor Company
Electronics Division
16800 Executive Plaza Drive
Regent Court Building
Suite 321
Dearborn, Michigan 48126

This paper describes inductor models developed by using generalized linear regression. Data for these models can be obtained by using either an impedance analyzer or a Q-meter. The resulting models, while based on low-frequency data, accurately predict inductor Q and impedance versus frequency and inductor self-resonant frequency.

Introduction

This paper is written to describe the application of statistical data reduction techniques to inductor modeling. The ultimate goal of this data reduction process is to produce models for inductors that accurately estimate observed measurements. The resulting models, obtained from closed-form equations, can be used as they are, or as starting points for optimizer-generated models. The inductor model to be used in this paper is shown in Figure 1.

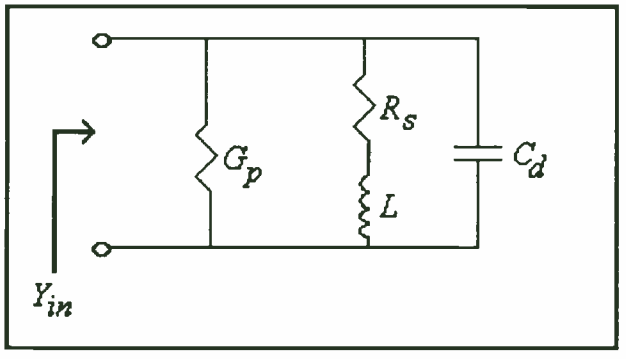


Figure 1. Four-element Inductor Model

The immediate goal of the modeling procedure is to find values for L, C_d, G_p, and R_s, based on laboratory measurements. It must be remembered that resulting models are *statistical* not *functional*. That is, since a finite-size physical structure is being modeled as a network of linear lumped elements, the models obtained are intended for use in the frequency and power range where the inductor to be modeled is linear and physically small relative to one wavelength. This has several implications. First among these is that estimates for self-resonant frequency based on low-frequency data may not agree with measured self-resonant frequency. This may be due to changes in core permeability as a function of frequency or transmission-line effects. Thus, attempts to reconcile observed oscillatory behavior with data obtained from these methods may prove frustrating. Viewed in another light, these methods produce data on distributed capacitance that may be unobtainable by measurement of self-resonant frequency.

A second implication is that these models will not work very well if the inductor is characterized at a low magnetic flux level, and then operated in saturation.

The paper is broken into sections covering background, general assumptions, lossless models, lossy models, and laboratory results.

Background

Inductances are fundamental lossless *elements* in many circuits, including filters[1], diplexers[2], directional couplers[3], matching networks[4], and switch-mode power supplies[5].

[1]Kuo, F. F. *Network Analysis and Synthesis.* Second edition (New York: Wiley, 1966), pp. 397-410

[2]Youla, D. C., Pillai, S. U., and Winter, F. "Theory and Design of Maximally Flat Low-Pass High-Pass Reactance Ladder Diplexers." *IEEE Transactions on Circuits and Systems-I*, Vol. 39 No. 5 (May 1992): pp. 337-349

[3]E. J. Tillo, "Design and Optimization of Lumped-Element LCM Directional Couplers," *RF Expo East Proceedings*, September 22-24, 1992, pp. 467-479

[4]Besser, L. "Designer Tips: Reactive Transformation of Resistances." *Applied Microwave Magazine* (Winter 1993): pp. 104-110

[5]Mitchell, D. M. *DC-DC Switching Regulator Analysis* (New York: McGraw-Hill, 1988), p. 7

Thus, in order to build such circuits it is necessary to fabricate *components*, called inductors, that approximate the desired inductance to some degree. For this to be accomplished, it is often necessary to obtain comprehensive and accurate measurements of inductor properties, which may include more than inductance. For purposes of this paper, losses and self-resonance will also be considered and accounted for with additional elements in the inductor model. Losses and self-resonance are a consequence of physical constraints, including capacitance between winding turns and winding resistance[6].

Several different instruments are suitable for measuring inductor properties, the two considered in this paper are the impedance analyzer and the Q-meter. With each of these instruments there is a temptation to obtain a simple inductor model by performing a measurement at a single frequency and then using the calibrated scales or displays to determine inductance without applying further data reduction techniques. These "single-frequency" methods may produce inductor models that have a great degree of inaccuracy. A brief explanation of single-frequency measurement procedures and the resulting models follows.

In the case of the impedance analyzer, a model consisting of an inductance and a single resistance can be readily obtained by merely connecting an inductor to the instrument terminals, entering a frequency value, and selecting a model for the inductor with a few keystrokes on the front panel of the instrument[7]. The available inductor models include an inductance in series with a resistance and an inductance in parallel with a conductance. A problem with this measurement technique is that the reactance of the inductor is due to both inductance and capacitance, and the resistance of the inductor is due to several frequency dependent phenomena. Thus, as the measurement frequency is varied, the inductive and resistive elements in the model also vary.

Inductor measurements are performed in a somewhat different fashion with the Q-meter. The routine procedure is to connect an inductor under test to the appropriate terminals on the Q-meter, forming a series resonant circuit with an internal oscillator and capacitor. The internal oscillator frequency is manually adjusted to a specified value f_0, and the internal capacitor is manually adjusted to produce a resonant condition. Resonance is indicated by a peak response on a voltmeter. When the peak is found the measured inductance and Q of the inductor under test, referred to as L_{qm} and Q_m, are read directly from calibrated scales on the capacitor and voltmeter respectively. Simple two element models can then be easily generated. For the case of the inductance with series loss, the resistor R_{sqm}, is found to be[8]:

$$R_{sqm} = \frac{\omega_0 L_{qm}}{Q_m}. \qquad (1)$$

For the case of parallel loss one obtains[9]:

$$G_{pqm} = \frac{1}{Q_m \omega_0 L_{qm}}. \qquad (2)$$

Where the angular frequency for both cases is defined by:

$$\omega_0 = 2\pi f_0. \qquad (3)$$

When this method is followed, L_{qm}, R_{sqm}, and G_{pqm}, vary with frequency.

In addition to the inconvenience of element values that vary with frequency, both of the single-frequency measurement techniques described above for the impedance analyzer and the Q-meter have a serious shortcoming in that they do not account for the self-resonance phenomenon.

General Assumptions

For the purposes of this paper three assumptions will be made at the start. First, the network shown in Figure 1 accurately models an inductor. Second, the condition of resonance will occur when the magnitudes of the reactances of the inductance L and capacitance C_d are equal. That is, the self-resonant frequency is given by:

$$f_{sr} = \frac{1}{2\pi\sqrt{LC_d}}. \qquad (4)$$

The third assumption is that the Q of an inductive circuit is defined by:

$$Q_{circuit} \equiv \frac{-\mathrm{Im}[Y_{in}]}{\mathrm{Re}[Y_{in}]}. \qquad (5)$$

In preparation of the application of Equation (5) to the circuit shown in Figure 1, the driving point admittance is written as:

$$Y_{in} = G_p + \frac{1}{R_s + j\omega L} + j\omega C_d. \qquad (6)$$

[6]Snelling, E. C. *Soft Ferrites: Properties and Applications.* Second edition (London: Butterworths, 1988), pp. 175-187

[7]Hewlett Packard, *Model 4192A LF Impedance Analyzer Operation and Service Manual* (Manual Part No. 04192-90001, March 1982), page 3-41

[8]ibid. page 3-41

[9]ibid. page 3-41

When this expression for Y_{in} is substituted into Equation (5) the result is:

$$Q_m = \frac{\frac{\omega L}{R_s^2 + \omega^2 L^2} - \omega C_d}{G_p + \frac{R_s}{R_s^2 + \omega^2 L^2}}. \qquad (7)$$

Where Q_m is the value that would be measured with either a Q-meter or impedance analyzer when the circuit is built with ideal elements. Hence, Equation (7) shows the measured value of Q for an inductor when the appropriate element values are used in the model.

When circuit losses are small, Equation (7) can be simplified by neglecting second-order loss terms:

$$Q_m = \frac{\frac{1}{\omega L} - \omega C_d}{G_p + \frac{R_s}{\omega^2 L^2}}. \qquad (8)$$

Equation (8) shows that the measured Q of an inductor will become lower as shunt capacitance increases. This shunt capacitance is the result of the so-called distributed capacitance of the inductor as well as any stray capacitance present at the terminals of the measuring instrument. Since it is beneficial to remove the effect of stray capacitance present due to the measuring instrument from the numerical value of Q, Equation (8) will be re-written with the capacitance C_d set equal to zero:

$$Q = \frac{\frac{1}{\omega L}}{G_p + \frac{R_s}{\omega^2 L^2}} \qquad (9)$$

The quantity Q will be referred to as the Q of the inductor. The m subscript has been dropped because this quantity cannot be directly measured.

Determination of Inductance and Distributed Capacitance

The process of obtaining estimates of the numerical values of C_d and L shown in Figure 1 will now be given. The crucial assumption at this point is that the losses in the coil are small enough to neglect. First, a data set is obtained by performing measurements on an inductor at a series of n different frequencies using either an impedance analyzer or Q-meter. Once this is done, generalized linear regression can be used to estimate the numerical values of C_d and L, and thus the self-resonant frequency of the inductor.

For the case of the impedance analyzer, the observed admittance will be considered to be a combination of an inductance and capacitance connected in parallel. The equation for the driving point admittance is:

$$\mathrm{Im}[Y_{in}(\omega)] = \omega C_d - \frac{1}{\omega L}. \qquad (10)$$

This equation can be re-written in the form used in the statistics literature[10]:

$$Y_i = B_0 + B_1 X_{i1}. \qquad (11)$$

Where the input admittance is a function of the angular frequency ω,

$$Y_i = \frac{\mathrm{Im}[Y_{in}(\omega)]}{\omega}; \omega = \omega_i, i = 1...n, \qquad (12)$$

the coefficient terms are capacitance,

$$B_0 = C_d, \qquad (13)$$

and reluctance,

$$B_1 = \frac{1}{L}, \qquad (14)$$

and the independent variable is a function of angular frequency:

$$X_{i1} = \frac{1}{\omega^2}; \omega = \omega_i, i = 1...n. \qquad (15)$$

At this point, the values of C_d and L can be found by using the methods of linear regression.

Capacitance and inductance estimates can also be made based on data obtained with Q-meter measurements. In this case the equations are slightly different, but the result is the same. The initial step is to write the condition of resonance at some frequency and Q-meter capacitor setting C:

$$\omega_0 = \frac{1}{\sqrt{L(C + C_d)}}. \qquad (16)$$

[10]Neter, J., Wasserman, W., and Kutner, M. H. *Applied Linear Regression Models* (Homewood, IL: Irwin, 1989), pp. 204-209

Equation (16) can be re-written as:

$$C = -C_d + \frac{1}{L\omega_0^2}. \qquad (17)$$

Equation (17) is of the form:

$$Y_i = B_0 + B_1 X_i, \qquad (18)$$

which is the form used in the statistics literature.

At this point, linear regression can be applied to Equation (17) to determine C_d and L. It should be noted that a similar technique has been applied to Q-meter data in the past. Determining C_d and L graphically from a plot of C versus $\frac{1}{\omega_o^2}$ is known as the "negative-intercept method"[11].

When either of the above procedures is followed, the estimates of inductance and distributed capacitance are constant as a function of frequency, and thus the values of C_d and L can be substituted into Equation (4) to find the self-resonant frequency of the inductor.

Determination of Series Resistance and Shunt Conductance

In order to account for inductor losses it is necessary to perform a series of measurements at n different frequencies. The resulting data can be used to obtain values of resistive elements in the inductor model. The technique will now be shown.

This technique is based on the value of Q for an inductor at a series of frequencies. This value of Q may be obtained from the Q-meter or a LF Impedance Analyzer with the application of a correction technique designed to remove the effect of shunt capacitance, C_d, from the Q measurement. The correction for Q-meter data, described in the *Model 4342A Q Meter Operating and Service Manual*[12], is:

$$Q = Q_m\left(\frac{C + C_d}{C}\right). \qquad (19)$$

Where, as before, C is the value indicated on the capacitance dial on the Q-meter.

The correction for the LF Impedance Analyzer data is similar, with the exception that the values used are the estimate of inductance, L, and the inductance, L_m, measured at a particular frequency:

$$Q = Q_m\left(\frac{L_m}{L}\right). \qquad (20)$$

The choice of which instrument to use depends on several factors, including the Q of the inductor to be modeled. When the inductor Q is above 100, the Q-Meter can be more accurate[13].

Equation (9), relating Q to the resistive and inductive elements for the circuit shown in Figure 1 can be re-written as:

$$\frac{1}{Q} = \frac{R_s}{\omega L} + \frac{\omega L}{R_p}. \qquad (21)$$

Where:

$$R_p \equiv \frac{1}{G_p}. \qquad (22)$$

Equation (21) can be re-written as:

$$\frac{\omega L}{Q} = R_s + \frac{\omega^2 L^2}{R_p}, \qquad (23)$$

which is of the form:

$$Y_i = B_0 + B_1 X_{i1}. \qquad (24)$$

Where Y_i is a function of the angular frequency ω,

$$Y_i = \frac{\omega L}{Q}; \omega = \omega_i, i = 1 \ldots n, \qquad (25)$$

[11]Rao, V. V. K., "The Q-Meter and Its Theory." *Proceedings of the I.R.E.*, Vol. 30 No. 11 (November 1942): pp. 502-505

[12]Hewlett Packard *Model 4342A Q Meter Operating and Service Manual* (Manual Part No. 04342-90009, August 1981), pages 3-15 to 3-16

[13]Honda, M. *The Impedance Measurement Handbook* (Hewlett Packard, 1989), pp. 5-4 to 5-7

the coefficient terms are resistance,

$$B_0 = R_s, \qquad (26)$$

and conductance,

$$B_1 = G_p, \qquad (27)$$

and the independent variable is a function of angular frequency,

$$X_{11} = \omega^2 L^2; \omega = \omega_i, i = 1 \ldots n. \qquad (28)$$

Linear regression can be applied to Equation (23) to obtain R_s and G_p.

Example

Consider the case of an air core inductor with a nominal inductance of 10 μH. Data sets obtained with a Q-meter and LF Impedance Analyzer are shown in Table 1 and 2 respectively. The measurement frequencies were selected to provide convenient settings of the Q-meter capacitance dial. The entries in the "single-frequency inductance" (*Lqm*) column in Table 1 were calculated by solving Equation (16) for the inductance L with the capacitance C_d set equal to zero.

F; MHz	*C*; pF	*Qm*	*Q*	*Lqm*; uH
2.3	470	210	213	10.14
3.33	220	239	247	10.3
4.65	110	257	273	10.63
6.68	50	261	297	11.37
8.90	25	236	301	12.81

Table 1. Q-Meter Data

F; MHz	*Gx*; uS	*Bx*; mS	*Qm*	*Q*	*Lm*; uH
2	55	-7.821	142	143	10.18
3	47	-5.166	110	112	10.27
4	39	-3.806	98	101	10.45
6	29	-2.389	82	90	11.10
9	24	-1.374	57	73	12.87

Table 2. LF Impedance Analyzer Data

There are a few observations to make about these data sets. First, the agreement of the single-frequency inductance *Lqm* in Table 1 and measured inductance *Lm* in Table 2 is quite good. This is to be expected with a low-loss inductor, where the major contribution to the impedance is due to inductive reactance. When losses are low the LF Impedance Analyzer makes an accurate estimate of the imaginary part of an unknown impedance and the Q-meter shows a sharp easy-to-locate peak near resonance. Consequently, precise measurements can be on high-Q inductors. The fact that the values of single-frequency and measured inductance vary as a function of frequency is due to the effect of shunt capacitance.

Second, note the large disagreement in inductor Q between Tables 1 and 2. No satisfactory explanation was found for this difference.

The data from Tables 1 and 2 were used to determine the elements for the model shown in Figure 1. The results are summarized in Table 3. Note that there is good agreement between inductance, capacitance, and self-resonant frequency values, but poor agreement with the resistance values. This is a consequence of the difference in measured Q values.

Element or parameter	Q-Meter	LF Imp. Ana.
L; uH	9.9949	10.0676
Cd; pF	6.9079	6.2441
Rs; ohms	.6731	.8749
Rp; k ohms	255.25	46.176
fsr; MHz	19.15	20.07

Table 3. Modeling Results

Calculated Q values are shown for the Q-meter and LF Impedance Analyzer results in Tables 4 and 5 respectively. The "model" values were obtained by substituting the estimates of element values into Equations (8) and (9) for Q_m and Q respectively. The Q-meter results show close agreement between measurement and model values, especially near the center frequency of the measurement range. The fit is not quite as good for the LF Impedance Analyzer results; this may be due to measurement error.

F; MHz	*Qm*; measured	*Qm*; from model	*Q*; calc. from meas.	*Q*; from model
2.3	210	188	213	191
3.33	239	240	247	248
4.65	257	273	273	289
6.68	261	271	297	307
8.90	236	231	301	294

Table 4. Q-Meter Results

F; MHz	*Qm*; measured	*Qm*; from model	*Q*; calc. from meas.	*Q*; from model
2	142	102	143	103
3	109	112	112	114
4	97	107	101	111
6	82	86	90	95
9	57	57	73	72

Table 5. LF Impedance Analyzer Results

Conclusion

Usefulness of data obtained from *Q*-meter and impedance analyzer measurements on inductors can be significantly increased with the application of simple data reduction techniques.

The circuit shown in Figure 1 with statistically derived element values produces an inductor model whose elements are not a function of frequency. This model also accounts for the observed self-resonance phenomenon as well as the change of inductor *Q* as a function of frequency.

The *Q*-meter may produce more accurate data on inductor *Q* than the impedance analyzer does, particularly when high-*Q* inductors are being measured.

Acknowledgment

The author is indebted to Ms. JoAnne Casey-Roney for her assistance with statistics texts, Mr. Steve Schwinke for his insightful discussions, and Professor A. B. Macnee for his suggestion of the negative-intercept method.

COMPUTER AIDED DESIGN TOOLS FOR SMALL SIGNAL RF MATCHING NETWORKS

BY

Michael Rothery, Sam Richie, and Madjid A. Belkerdid

University of Central Florida
Electrical and Computer Engineering Department

Orlando, Florida 32816

ABSTRACT

This paper presents PC based computer aided design tools for small signal RF amplifiers and oscillators in the VHF and UHF bands. The development of these systems is based on the small signal S parameters and noise parameters which can supply information on stability, gain, and noise figure. The manual work generally involved in developing solutions for these types of problems can become tedious and time consuming.

The input to the CAD system consists of the operating frequency, the four small signal S parameters, the noise figure parameters, and the system resistance. The CAD system then plots the stability circles, the constant gain circles, and the constant noise figure circles on a graphical display of the Smith chart.

A user based iterative optimization routine may then be employed to complete the hardware requirements for the matching network design. This design method provides for a user friendly, interactive, fast, and inexpensive PC based design tool.

I. INTRODUCTION

This paper first demonstrates the computerized algorithm used to determine the component values of a matching network. For the given reflection coefficient on the Smith chart, an intersection of circles technique is used to determine the amount of reactance or susceptance required from each component in the matching network. Then, given the frequency of operation and the system resistance, the value of the inductor or capacitor may then be determined.

Some of the basic theory that is used by the CAD system for designing two-port small signal amplifiers is then reviewed. The design concepts are based on the information provided by the S parameters and noise parameters of a biased active device. Used in conjunction with the Smith chart, these parameters supply the information necessary to develop the proper matching networks for the input and output ports. The matching networks determine the stability, gain, and noise figure of the small signal, two-port amplifier.

Finally, the method employed by the CAD system for the design of a negative resistance oscillator will be presented. Instead of matching the input and output ports of the active device, a resonant circuit is chosen for one port, while the other is matched under loaded conditions.

II. COMPUTERIZED SMITH CHART

A computerized Smith chart is the heart of the CAD tools. A mathematical model is first presented, followed by the Smith chart concept, and a design example.

A. Mathematical Model

The Smith chart may be used to transform a reflection coefficient, Γ, to a normalized impedance, z, following the relation [1]

$$z = \frac{1+\Gamma}{1-\Gamma}. \qquad (1)$$

The reflection coefficient may be expressed in polar coordinates or rectangular coordinates. In polar form, $\Gamma = m\angle\theta$, where $-180 \le \theta \le 180$ and for any passive reflection coefficient, $m < 1$. In rectangular form, Γ is given as $\Gamma = U + jV$, where of course $U = m\cos(\theta)$ and $V = m\sin(\theta)$.

The normalized impedance, z, will be given in the form $z = z_R + jz_X$, where z_R and z_X represent the normalized resistance and normalized reactance components, respectively. The normalized admittance is given by $y = y_G + jy_B$, where y_G is the normalized conductance and y_B is the normalized susceptance.

Substituting $z = z_R + jz_X$ and $\Gamma = U + jV$ into (1) yields the relationship

$$z_R + jz_X = \frac{(1+U)+jV}{(1-U)-jV}.$$

Separating this expression into its real and imaginary parts yields

$$z_R = \frac{1-U^2-V^2}{(1-U)^2+V^2}$$

and

$$z_X = \frac{2V}{(1-U)^2+V^2}. \qquad (2)$$

Given the location of Γ, the expressions for z_R and z_X respectively provide the values of the normalized resistance and normalized reactance circles in the Smith chart.

In order for the CAD system to determine the component values of a matching network, it is necessary to know the amount of reactance or susceptance needed from each component. The Smith chart naturally lends itself to this. In order to "read" the reactance or susceptance from the Smith chart, an intersection point between two circles will be calculated. The following derivation is for the value of the reactance circle at the intersection point between the normalized unit constant conductance circle (y_G=1) and a normalized constant resistance circle where $0 \le z_R \le 1$.

Assume the constant unit conductance circle is centered at the rectangular coordinates (U_1 , 0) with radius r_1 as shown in Figure 1. Let the point to match be selected on a constant resistance circle centered at (U_2 , 0) with radius r_2 .

The equations for these two circles are

$$(U-U_1)^2+(V-0)^2 = r_1^2$$
$$(U-U_2)^2+(V-0)^2 = r_2^2.$$

This system of equations may be written

$$U^2 - 2\cdot U\cdot U_1 + U_1^2 + V^2 = r_1^2 \qquad (3)$$
$$U^2 - 2\cdot U\cdot U_2 + U_2^2 + V^2 = r_2^2 \qquad (4)$$

Subtracting (4) from (3) yields

$$-2\cdot U\cdot U_1 + 2\cdot U\cdot U_2 + U_1^2 - U_2^2 = r_1^2 - r_2^2, \qquad (5)$$

where U now represents the horizontal offset from the center of the Smith chart to where the two circles intersect.

Substituting U_{int} for U and noting that U_1 = -.5 and r_1 = .5, equation (5) may be written

$$U_{int} = \frac{r_1^2-r_2^2-U_1^2+U_2^2}{2\cdot U_2-2\cdot U_1} = \frac{.5^2-r_2^2-.5^2+U_2^2}{2\cdot U_2+1}$$

$$= \frac{U_2^2-r_2^2}{2\cdot U_2+1}. \qquad (6)$$

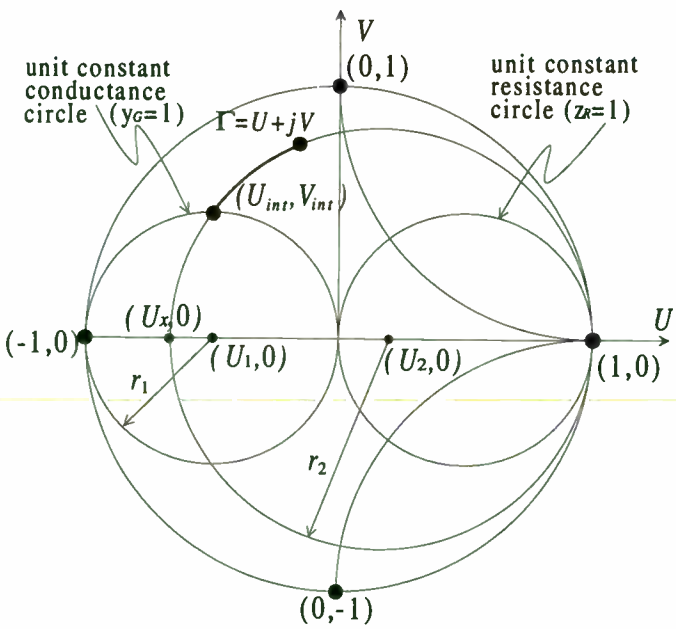


Figure 1. Intersection of constant resistance circle with unit constant conductance circle

To determine U_2 and r_2 , we first need to find the point where the resistance circle crosses the real (V=0) axis. Note that equation (1) may be rewritten in the form

$$\Gamma = \frac{z-1}{z+1},$$

which leads to the relationship

$$U + jV = \frac{(z_R-1)+jz_X}{(z_R+1)+jz_X}.$$

Separating this into its real and imaginary parts yields

$$U = \frac{z_R^2 - 1 + z_X^2}{(z_R+1)^2 + z_X^2}$$

and

$$V = \frac{2z_X}{(z_R+1)^2 + z_X^2}.$$

Along the real axis, the solution for U leads to the result

$$U = \frac{z_R - 1}{z_R + 1}\Bigg|_{z_X = 0}.$$

This expression may be generalized to the form

$$U_x = \frac{z_R - 1}{z_R + 1},$$

where $(U_x, 0)$ is the point where the given constant resistance circle, z_R, crosses the real axis of the Smith chart. Since all constant resistance circles also pass through the point $(U, V) = (1, 0)$, the radius, r_2 , of a given constant resistance circle may be found as half the distance between points (1, 0) and $(U_x, 0)$; or

$$r_2 = \frac{1 - U_x}{2} = \frac{1 - \frac{z_R - 1}{z_R + 1}}{2} = \frac{1}{z_R + 1}. \qquad (7)$$

U_2 may then be found as the relation

$$U_2 = U_x + r_2 = \frac{z_R}{z_R + 1}. \qquad (8)$$

Substituting (7) and (8) into (6) gives the horizontal offset of the intersection point from the Smith chart center as a function of only z_R.

$$U_{int} = \frac{z_R - 1}{3z_R + 1}, \quad 0 \le z_R \le 1. \qquad (9)$$

To find the vertical offset of the intersection point from the center of the Smith chart, simply substitute U_{int} for U and V_{int} for V in equation (3) and solve for V_{int}, yielding

$$V_{int} = \pm\sqrt{-U_{int}^2 - U_{int}}, \quad -1 \le U_{int} \le 0$$

$$= \frac{\pm 2\sqrt{z_R - z_R^2}}{3z_R + 1}, \quad 0 \le z_R \le 1. \qquad (10)$$

From (9) and (10), it can be seen that the intersection points $(U_{int}, \pm V_{int})$ may be obtained as functions of only the constant resistance circle, z_R.

It is then possible to directly find the value of the normalized reactance at the intersection point given any normalized resistance circle $0 \le z_R \le 1$. By substituting U_{int} and V_{int} for U and V in equation (2), the normalized reactance at the intersection point may be found directly as a function of z_R by

$$z_{X,int} = z_X = \pm\sqrt{z_R - z_R^2}, \quad 0 \le z_R \le 1 \qquad (11)$$

The CAD system uses the same geometric layout for an admittance chart as it does for the impedance chart just described. Therefore, the determination of the normalized susceptance at the intersection of a constant conductance circle $0 \le y_G \le 1$ and the unit resistance circle can easily be shown to be

$$y_{B,int} = \pm\sqrt{y_G - y_G^2}, \quad 0 \le y_G \le 1 \qquad (12)$$

B. Matching network design using the computerized Smith chart

The following steps outline the procedure used by the CAD system for developing four two-element matching networks for a given impedance using the Smith chart.

1) Assume the point is selected from the Smith chart at $\Gamma_1 = U_1 + jV_1$ and has the associated impedance $z_1 = z_{R1} + jz_{X1}$, where $z_{R1} < 1$.
2) Equation (11) provides the value of the normalized reactance circle at the intersection point between the constant resistance circle, z_{R1} , and the upper half of the unit constant conductance circle (i.e. the positive solution for $z_{X,int}$). This will be $z_{int} = z_{R,int} + jz_{X,int}$, where $z_{R,int} = z_{R1}$.
3) Calculate the admittance associated with z_{int}. This yields $y_{int} = 1/z_{int} = y_{G,int} + jy_{B,int}$, where $y_{G,int}$ will always be unity because y_{int} lies on the normalized unit constant conductance circle.
4) The normalized reactance needed from the series element of this matching network is given by $z_{X,int} - z_{X1}$; and the normalized susceptance needed from the shunt element is $0 - y_{B,int}$.
5) Repeat step 2, but use the negative solution for $z_{X,int}$. Repeat steps 3 and 4.
6) Let y_1 be the admittance associated with z_1. ($y_1 = 1/z_1 = y_{G1} + jy_{B1}$, where $y_{G1} < 1$).

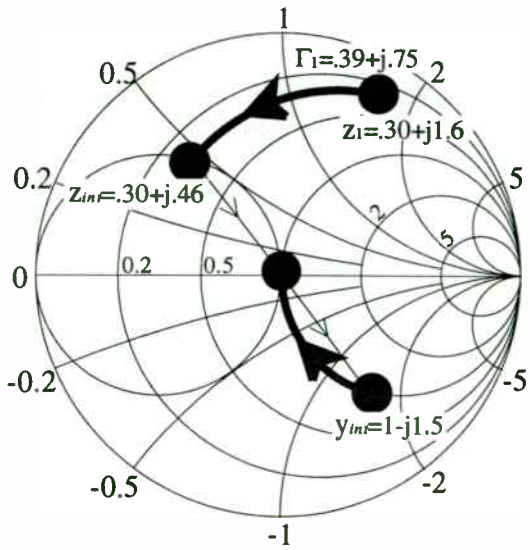


Fig. 2a

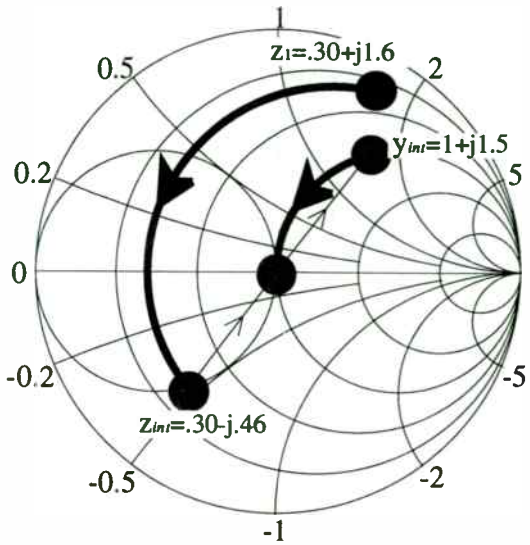


Fig. 2b

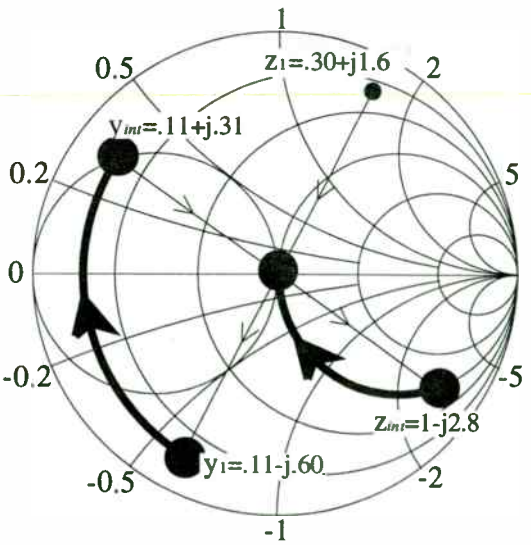


Fig. 2c

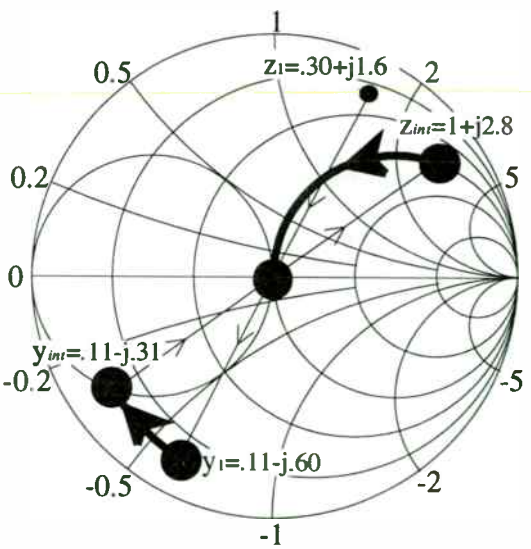


Fig. 2d

Figure 2. Methods for matching a given impedance

7) Using equation (12), solve for the positive solution of $y_{B,int}$ to obtain $y_{int} = y_{G,int} + jy_{B,int}$, where $y_{G,int} = y_{G1}$.
8) Calculate the impedance of y_{int} to obtain $z_{int} = 1/y_{int} = z_{R,int} + jz_{X,int}$, where $z_{R,int}=1$.
9) The susceptance needed from the shunt element will be $y_{B,int} - y_{B1}$; and the reactance needed from the series element is $0 - z_{X,int}$.
10) Repeat step 7, but use the negative solution for $y_{B,int}$. Repeat steps 8 and 9.

It should be noted that if a point to be matched is selected from the impedance Smith chart within the unit constant conductance circle ($y_{G1} > 1$) , only steps 1 through 5 may be used for developing a two-element matching network. On the other hand, if the point is selected within the unit constant resistance circle ($z_{R1} > 1$), then only steps 6 through 10 apply.

C. Design example

An example implementation of the design procedure outlined above is described in the following steps and graphically depicted in Figure 2.

1) Assume the point selected is at the position $\Gamma_1 = U_1 + jV_1 = .39 + j.75$ as shown in Figure 2a. Using equation (1), the associated impedance is found to be $z_1 = z_{R1} + jz_{X1} = 0.30 + j1.6$.
2) Equation (11) determines the reactance at the intersection point between the unit constant conductance circle and the constant resistance circle, $z_{R1} = 0.30$; this yields the point $z_{int} = z_{R,int} + jz_{X,int} = 0.30 + j.46$.
3) The admittance associated with z_{int} is $y_{int} = 1/z_{int} = y_{G,int} + jy_{B,int} = 1.0 - j1.5$.
4) A series element with reactance $j(z_{X,int} - z_{X1}) = j(0.46\text{-}1.6) = -j1.14$, and a shunt element with susceptance $j(0 - (-1.5)) = +j1.5$ are required for the matching network.
5) Using the negative solution for equation (11) provides the following information: (see Fig. 2b)
$z_{int} = .30 - j.46$; $y_{int} = 1.0 + j1.5$.
series element: $-j2.06$ reactance.
shunt element: $-j1.5$ susceptance.
6) $y_1 = 1/z_1 = .11 - j.60$. (see Fig. 2c)
7) Calculating $y_{B,int}$ from (12) leads to $y_{int} = .11 + j.31$.
8) $z_{int} = 1.0 - j2.8$.
9) shunt element: $j(.31\text{-}(-.61)) = +j.91$ susceptance.
series element: $+j2.8$ reactance
10) $y_{int} = .11 - j.31$; $z_{int} = 1.0 + j2.8$. (see Fig. 2d)
shunt element: $(j(-.31\text{-}(-.60)) = +j.29$ susceptance.
series element: $-j2.8$ reactance.

Given the amount of reactance or susceptance needed, the value of the necessary component may then be calculated. The capacitive reactance, x_c, and the inductive reactance, x_L, are provided by the impedance chart. The capacitive reactance is given by

$$jx_c = \frac{1}{jR_0(2\Pi f_0 C)},$$

where, R_0 = system resistance,
f_0 = frequency of operation,
and C = capacitor value.

The capacitor value is then easily shown to be

$$C = -\frac{1}{R_0(2\pi f_0 x_c)}, \quad x_c < 0. \tag{13}$$

Similarly, the normalized reactance of an inductor may be found by

$$jx_L = \frac{j2\pi f_0 L}{R_0}, \text{ where } L \text{ is the inductor value.}$$

Solving for L yields

$$L = \frac{x_L R_0}{2\pi f_0}, \quad x_L > 0. \tag{14}$$

The capacitive susceptance, b_c, and the inductive susceptance, b_L, are read from the admittance chart. The capacitor and inductor values are then given by

$$C = \frac{b_c}{R_0(2\pi f_0)}, \quad b_c > 0 \tag{15}$$

and

$$L = -\frac{R_0}{(2\pi f_0) b_L}, \quad b_L < 0. \tag{16}$$

By knowing which chart, impedance or admittance, is being "read" from, the component values for the matching network may be calculated using equations (13) through (16).

III. SMALL SIGNAL AMPLIFIER DESIGN

Some of the basic principles applied to the design of transistor amplifiers include stability, gain, and noise analysis. The method of design employed by the CAD system is based on the S parameters of the DC biased transistor circuit at the frequency of operation. The S parameters provide all the information necessary to design for the desired stability and gain. In addition,

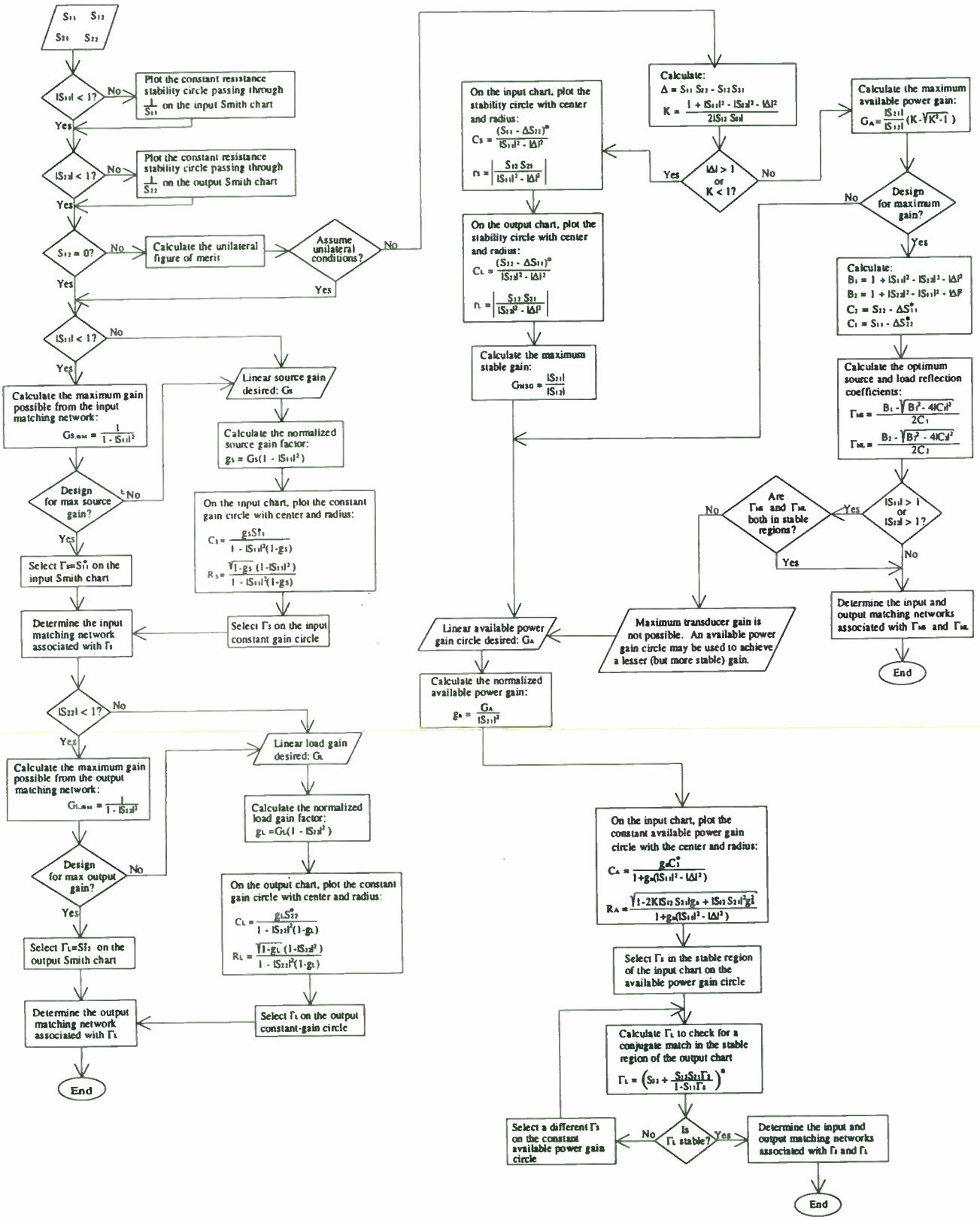


Figure 3. Flowchart for amplifier design

given the noise parameters of the circuit, the user may factor the desired noise figure criteria into the design.

The flow chart in Figure 3 provides for the manual development of stability circles and constant gain circles under unilateral (S_{12}=0) and bilateral conditions.

A. Stability analysis

Referring to the single stage amplifier model of Figure 4, a requirement for a two-port network to be considered unconditionally stable at a given frequency is

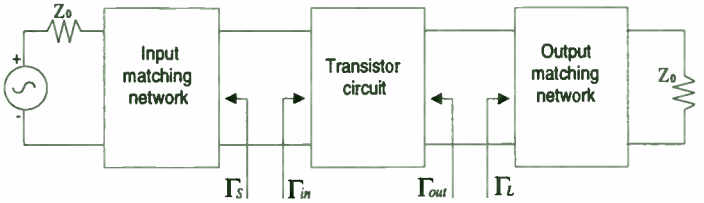


Figure 4. Two-port amplifier block diagram

for both the input and output port impedances to produce a positive real part. This requirement implies that $|\Gamma_{in}| < 1$ and $|\Gamma_{out}| < 1$, where Γ_{in} and Γ_{out} are defined as [1]

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L} \qquad (17)$$

and

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1-S_{11}\Gamma_S}. \qquad (18)$$

Note that Γ_S and Γ_L are the source and load reflection coefficients to be selected from the Smith chart Γ-plane. Since Γ_{in} is a function of Γ_L, and Γ_{out} is a function of Γ_S, it is true that the stability of the amplifier depends upon the matching networks to be determined by the selected locations of Γ_S and Γ_L. Therefore, it is necessary to determine the loci of Γ_S which produce $|\Gamma_{out}| = 1$ and the loci of Γ_L which produce $|\Gamma_{in}| = 1$. These values may be shown to lie in circles called stability circles whose centers and radii are functions of the S parameters. There has been a wealth of previously published work [1,2] regarding the derivation of the stability circles for the Smith chart Γ_S and Γ_L planes, only the results of which are presented here. However, it is first helpful to know whether or not the development of the stability circles is even necessary. A device may be determined to be unconditionally stable without resorting to the task of calculating the location of the stability circles. First, define Δ (the determinant of the S parameter matrix) as

$$\Delta = S_{11}S_{22} - S_{12}S_{21}.$$

The necessary and sufficient requirements for unconditional stability when $|S_{11}| < 1$ and $|S_{22}| < 1$ may be shown to be [3,4]

$$K = \frac{1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2}{2|S_{12}S_{21}|} > 1$$

and $|\Delta| < 1$.

Satisfying the above criteria indicates that the stability circles do not enter the passive region of the Smith chart Γ_S-plane or Γ_L-plane; therefore their calculation and plot is unnecessary.

However, under conditions where a device is determined to be potentially unstable (i.e. K<1 or $|\Delta| > 1$), it becomes necessary to plot the input and output stability circles as follows:

In the Γ_S-plane, the *input stability circle* is centered at C_S with radius R_S, where

$$C_S = \frac{(S_{11}-\Delta S_{22}^*)^*}{|S_{11}|^2-|\Delta|^2}$$

and

$$R_S = \left|\frac{S_{12}S_{21}}{|S_{11}|^2-|\Delta|^2}\right|.$$

In the Γ_L-plane, the *output stability circle* is centered at C_L with radius R_L, where

$$C_L = \frac{(S_{22}-\Delta S_{11}^*)^*}{|S_{22}|^2-|\Delta|^2}$$

and

$$R_L = \left|\frac{S_{12}S_{21}}{|S_{22}|^2-|\Delta|^2}\right|.$$

In short, the input stability circle represents the boundary along which $|\Gamma_{out}| = 1$, and the output stability circle represents the boundary along which $|\Gamma_{in}| = 1$.

It is then necessary to determine which values of Γ_S (inside or outside the stability circle) are in the stable region ($|\Gamma_{out}| < 1$) of the input chart, and which values of Γ_L are within the stable region ($|\Gamma_{in}| < 1$) of the output chart. Referring to the definition for Γ_{in} above, if $\Gamma_L = 0$ (the center of the output Smith chart), then $|\Gamma_{in}|$ is simply $|S_{11}|$. Therefore, if $|S_{11}| < 1$ then $|\Gamma_{in}| < 1$ and the center of the output chart is a point in the stable region. Conversely, if $|S_{11}| > 1$ then $|\Gamma_{in}| > 1$ and the center is a point in the unstable region. Similarly, if Γ_S=0 and $|S_{22}| < 1$ then $|\Gamma_{out}| < 1$ and the center of the input chart is in the stable region, and if $|S_{22}| > 1$ then $|\Gamma_{out}| > 1$ and the input chart center is part of the

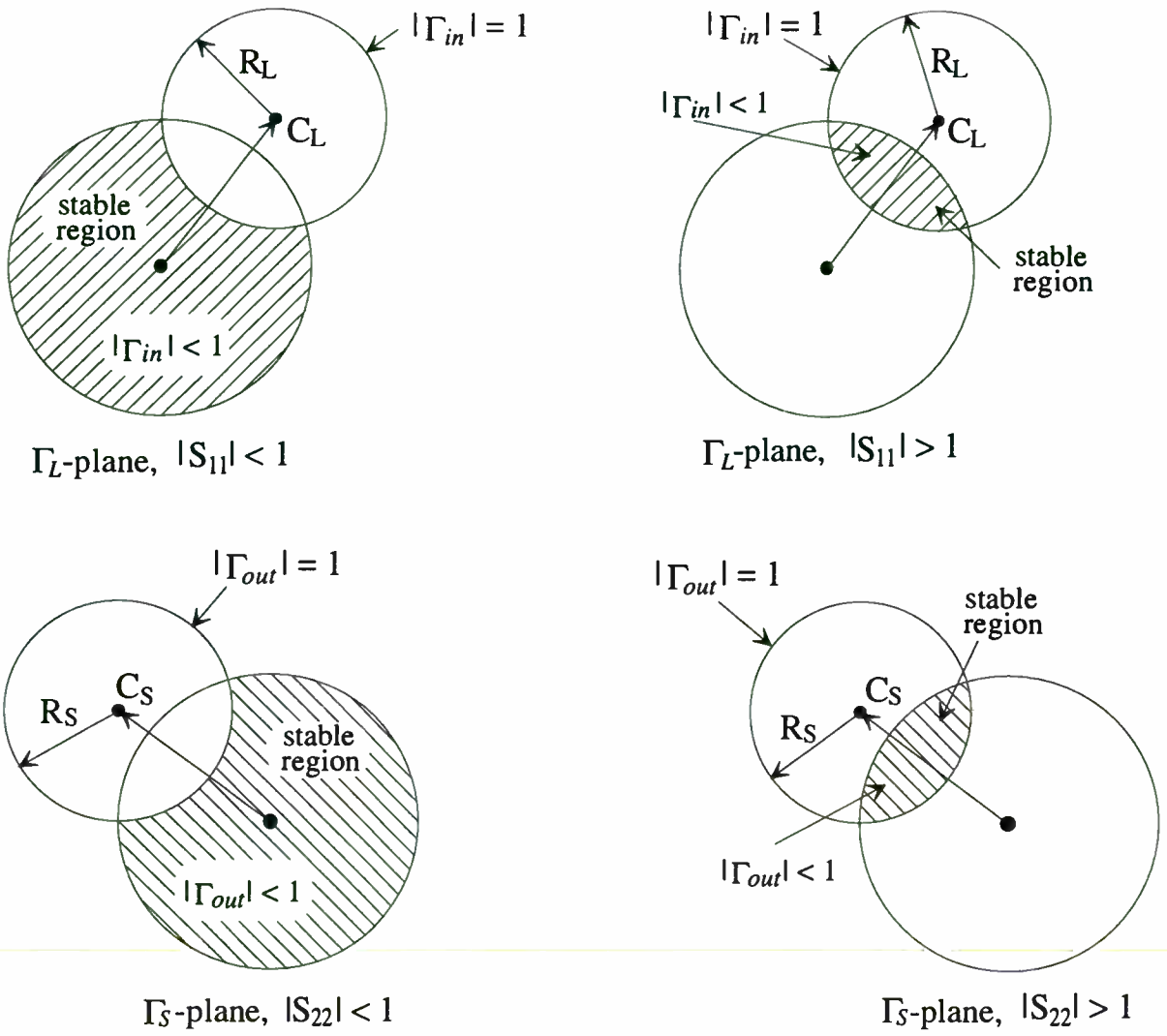


Figure 5. Stability circles for $|\Gamma_{in}|=1$ and $|\Gamma_{out}|=1$

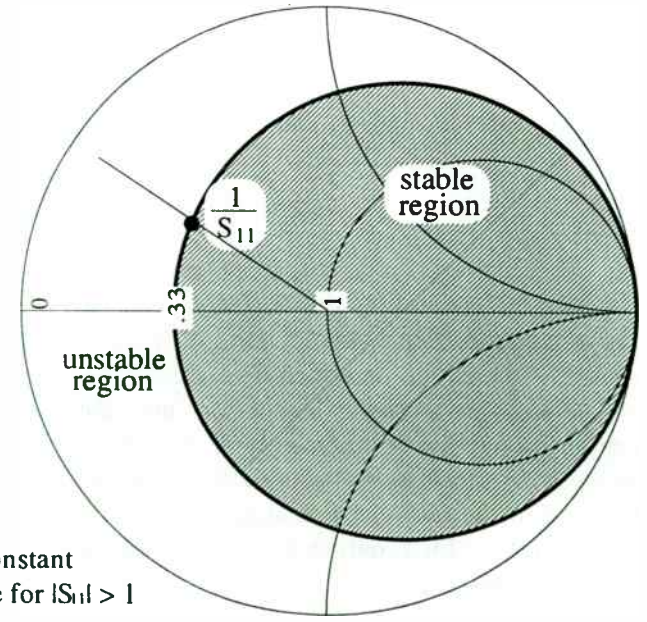


Figure 6. Example of constant resistance stability circle for $|S_{11}| > 1$

unstable region. These results may be described graphically as in Figure 5.

A second requirement for unconditional stability on the Smith chart is for $|S_{11}| < 1$ and $|S_{22}| < 1$. For instance, given an input reflection coefficient, S_{11}, whose magnitude is greater than unity implies that the real part of the impedance associated with S_{11} is negative and there is a potential for instability. For this reason, the real part of the source impedance, Z_S, associated with the selected source reflection coefficient, Γ_S, must provide for a total input loop resistance that is positive.

The minimum normalized resistance that Γ_S can acquire may be determined by plotting the critical point $\Gamma_{S,C} = \frac{1}{S_{11}}$. Γ_S must then be selected such that the real part of the impedance associated with Γ_S is greater than the real part of the impedance associated with $\Gamma_{S,C}$. In other words, $\mathrm{Re}(Z_S) > \mathrm{Re}(Z_{S,C})$. A similar argument may be followed if $|S_{22}| > 1$. The minimum normalized resistance required from Γ_L in the output chart is determined from the critical point $\Gamma_{L,C} = \frac{1}{S_{22}}$.

A short example will clarify the situation. An input reflection coefficient $S_{11} = 1.77\angle -130$ has an associated normalized impedance $Z_{S_{11}} = -.33 - j.42$. The critical point $\Gamma_{S,C} = \frac{1}{S_{11}} = .565\angle 130$ is calculated and found from the Smith chart to have the associated normalized impedance $Z_{S,C} = .33 + j.42$. Therefore, the source reflection coefficient, Γ_S, must be selected such that its normalized real part is within the +.33 normalized constant resistance circle on the Smith chart as shown in Figure 6. In addition, it will be shown that an infinite value for the gain is calculated by selecting the reflection coefficient at the critical point.

B. Transducer Gain Analysis

The CAD system follows two separate procedures for the calculation of constant gain circles, one for the unilateral case ($S_{12} = 0$) and the second for the bilateral case. Under unilateral conditions the gain may be expressed in the form

$$G_S = \frac{1-|\Gamma_S|^2}{|1-S_{11}\Gamma_S|^2}$$

and

$$G_L = \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2},$$

where G_S and G_L represent the gain produced by the input and output matching networks, respectively.

As stated earlier, if $|S_{11}|$ or $|S_{22}|$ is greater than unity, then by substituting $\Gamma_S = \Gamma_{S,C} = \frac{1}{S_{11}}$ or $\Gamma_L = \Gamma_{L,C} = \frac{1}{S_{22}}$ into the expressions for G_S or G_L, respectively, an infinite value for that particular gain will be calculated.

Under unilateral and unconditionally stable conditions the maximum values of G_S and G_L may be obtained by letting $\Gamma_S = S_{11}^*$ and $\Gamma_L = S_{22}^*$. The expressions may then be written

$$G_{S\max} = \frac{1}{1-|S_{11}|^2}$$

and

$$G_{L\max} = \frac{1}{1-|S_{22}|^2}.$$

However, a unilateral transducer gain other than $G_{S\max}$ or $G_{L\max}$ may sometimes be desired. The values for which Γ_S or Γ_L produce values of $G_S < G_{S\max}$ or $G_L < G_{L\max}$, respectively, may be shown [1,2] to lie in circles on a Smith chart. The center and the radius of the *input constant-gain circles* for the unilateral case are given by

$$C_S = \frac{g_S S_{11}^*}{1-|S_{11}|^2(1-g_S)}$$

and

$$R_S = \frac{\sqrt{1-g_S}\,(1-|S_{11}|^2)}{1-|S_{11}|^2(1-g_S)}.$$

where $g_S = \frac{G_S}{G_{S\max}}$ is known as the normalized input gain factor with G_S as the desired gain produced by the input network.

The center and radius of the *output constant-gain circles* under unilateral conditions can be shown to be

$$C_L = \frac{g_L S_{22}^*}{1-|S_{22}|^2(1-g_L)}$$

and

$$R_L = \frac{\sqrt{1-g_L}\,(1-|S_{22}|^2)}{1-|S_{22}|^2(1-g_L)}$$

where $g_L = \frac{G_L}{G_{L\max}}$ is the normalized output gain factor with G_L as the desired gain produced by the output network.

Under bilateral conditions, the CAD system provides two options to the user for the design of a desired transducer power gain. The first is to select Γ_S

from an available power gain circle, G_A , in the stable region on the input chart, then to calculate and plot its conjugate match (i.e. $\Gamma_L = \Gamma^*_{out}$) on the output chart to determine the stability for an output matching network. If Γ_L does not provide for the proper stability, then another point may be selected on the desired available power gain circle until the appropriate stability is achieved. The second option is similar except that the first point is selected from an operating power gain circle, G_P , in the Γ_L-plane, and the conjugate match (i.e. $\Gamma_S = \Gamma^*_{in}$) is then calculated and plotted on the input chart.

If a design procedure using an available power gain circle is to be followed, the desired available power gain circle, G_A , may be shown to lie in the Γ_S-plane centered at C_A with radius R_A given by

$$C_A = \frac{g_a C_1^*}{1+g_a(|S_{11}|^2-|\Delta|^2)}$$

and

$$R_A = \frac{\sqrt{1-2K|S_{12}S_{21}|g_a+|S_{12}S_{21}|^2 g_a^2}}{1+g_a(|S_{11}|^2-|\Delta|^2)}$$

where $g_a = \frac{G_A}{|S_{21}|^2}$

and $C_1 = S_{11} - \Delta S_{22}^*$.

On the other hand, if an operating power gain circle, G_P , is to be selected from the Γ_L-plane, then the center, C_P , and radius, R_P , are given by

$$C_P = \frac{g_p C_2^*}{1+g_p(|S_{22}|^2-|\Delta|^2)}$$

and

$$R_P = \frac{\sqrt{1-2K|S_{12}S_{21}|g_p+|S_{12}S_{21}|^2 g_p^2}}{1+g_p(|S_{22}|^2-|\Delta|^2)}$$

where $g_p = \frac{G_P}{|S_{21}|^2}$

and $C_2 = S_{22} - \Delta S_{11}^*$.

When the S parameters of the device indicate unconditionally stable bilateral conditions, the maximum available power gain or the maximum operating power gain is obtained at the point where R_A=0 or R_P=0. Either case leads to the following expression for maximum gain [1]

$$G_{max} = \frac{|S_{21}|}{|S_{12}|}(K - \sqrt{K^2-1}).$$

Under potentially unstable conditions where K < 1, the maximum stable gain is given by

$$G_{MSG} = \frac{|S_{21}|}{|S_{21}|}.$$

G_{MSG} is the gain that can be achieved by resistively loading the transistor to make K = 1 and then simultaneously conjugately matching [1] the input and output ports. Any time a simultaneous conjugate match is performed, it is assumed that the active device is unconditionally stable. If K < 1, oscillations may occur if either of the power gains are selected such that $G_A > G_{MSG}$ or $G_P > G_{MSG}$.

C. Noise figure analysis

In addition to stability and gain, another consideration in amplifier design is its noise figure. The noise figure of a two-port amplifier may be given in the form [2]

$$F = F_{min} + \frac{4R_N}{Z_0}\frac{|\Gamma_S-\Gamma_{opt}|^2}{(1-|\Gamma_S|^2)|1+\Gamma_{opt}|^2},$$

where

R_N = equivalent noise resistance of the transistor.

Γ_S = source reflection coefficient on the input chart.

Γ_{opt} = optimum reflection coefficient resulting in minimum noise figure.

F_{min} = minimum noise figure of the transistor.

Z_0 = system impedance.

This expression defines a family of noise figure circles in the Γ_S-plane. To compute the center and radius of a noise figure circle, it is first necessary to calculate the noise figure parameter, *N*, defined as

$$N = \frac{F-F_{min}}{4R_N/Z_0}|1+\Gamma_{opt}|^2.$$

The center, C_F , and radius, R_F , of a constant noise figure circle are then given by

$$C_F = \frac{\Gamma_{opt}}{N+1}$$

and

$$R_F = \frac{\sqrt{N(N+1-|\Gamma_{opt}|^2)}}{N+1}.$$

It is generally not possible to obtain a minimum noise figure and a maximum gain in the same design, this would only be possible if $\Gamma_{opt} = \Gamma_{in}^*$. Therefore, a compromise between the two may be reached by plotting both the constant gain circles and the constant noise figure circles on the input chart and selecting Γ_S to obtain an acceptable trade-off.

IV. Oscillator Design

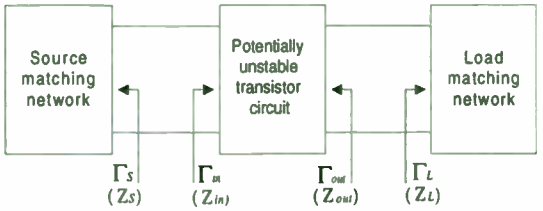


Figure 7. Two-port transistor oscillator model

The same S parameters used in the design of a two-port amplifier may be applied to the design of a two-port oscillator. One principal difference, however, is that the oscillator ports are passively terminated such that the biased transistor circuit is operating in an unstable region. This implies that the unmatched device should be potentially unstable (i.e. K<1 or $|\Delta|>1$) to function in an oscillator circuit.

To be consistent, the same reflection coefficient notation will be used for the oscillator design as for the amplifier design. Referring to the oscillator block diagram in Figure 7, further necessary conditions for oscillation can be expressed as

$$\Gamma_{in}\Gamma_S = 1 \qquad (19)$$

and

$$\Gamma_{out}\Gamma_L = 1. \qquad (20)$$

It can be shown [3] that if (19) is satisfied then (20) will also be satisfied, and vice versa.

If an active device is potentially unstable, the following steps will outline the procedure to meet the conditions given in equations (19) and (20).

1) Select Γ_S at any point in the unstable region indicated by the stability circle on the input Smith chart.
2) Calculate Γ_{out} as given in equation (18)
3) Γ_L may then be located from equation (20) as

$$\Gamma_L = \frac{1}{\Gamma_{out}}.$$

4) Determine the matching networks associated with Γ_S and Γ_L.

For completeness, the design for calculating Γ_S from a selected Γ_L is given.

1) Select Γ_L at any point in the unstable region indicated by the stability circle on the output Smith chart.
2) Calculate Γ_{in} as given in equation (17)
3) The location of Γ_S from equation (19) is

$$\Gamma_S = \frac{1}{\Gamma_{in}}.$$

4) Determine the matching networks associated with Γ_S and Γ_L.

V. Computer Aided Design Summary

The input to the CAD system consists of the operating frequency, the four small signal S parameters, the system resistance, and the noise figure parameters. The opening screen displays two Smith charts representing the Γ_S and Γ_L planes (i.e. the input and output charts, respectively). However, each chart may also be displayed individually in a larger scale if desired. If the S parameters indicate potential instability, the input and output stability circles will automatically be plotted.

From 'pull-down' menus the user is then able to plot constant gain circles in both the Γ_S and Γ_L planes. In addition, constant noise figure circles may be added to the Γ_S plane . A user based iterative optimization routine may then be used to design the hardware requirements for the matching network.

The optimization routine allows the user to continuously move a pointer around the computer generated image of the Smith chart. This permits selection of a position for the reflection coefficient that meets the requirements for a specified design. On this same Smith chart may be displayed the stability circle, constant gain circles, or constant noise figure circles. As the reflection coefficient is being determined, the possible networks to achieve the match for the port of interest are displayed. This method provides the user with immediate information on stability, gain, noise figure, and component values. Sample screens of the computerized Smith chart are shown in Figures 8, 9, 10, and 11.

REFERENCES

[1] G. Gonzales, *Microwave Transistor Amplifiers Analysis and Design*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[2] D.M. Pozar, *Microwave Engineering*, Addison-Wesley, Reading, Mass, 1990.

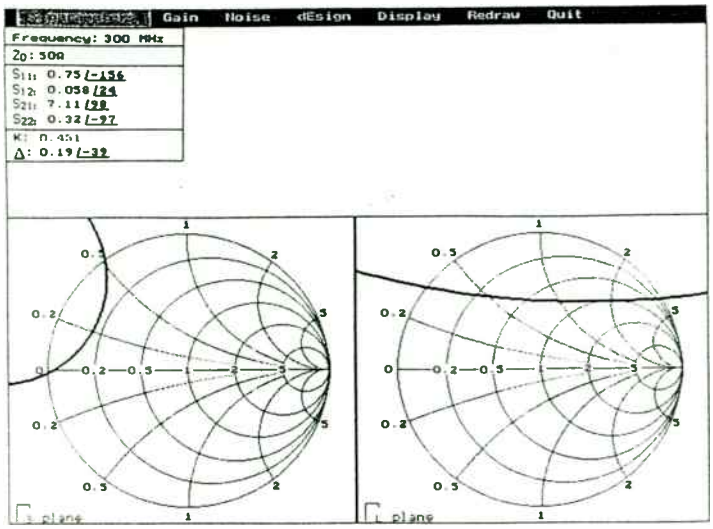


Figure 8. Stability circles

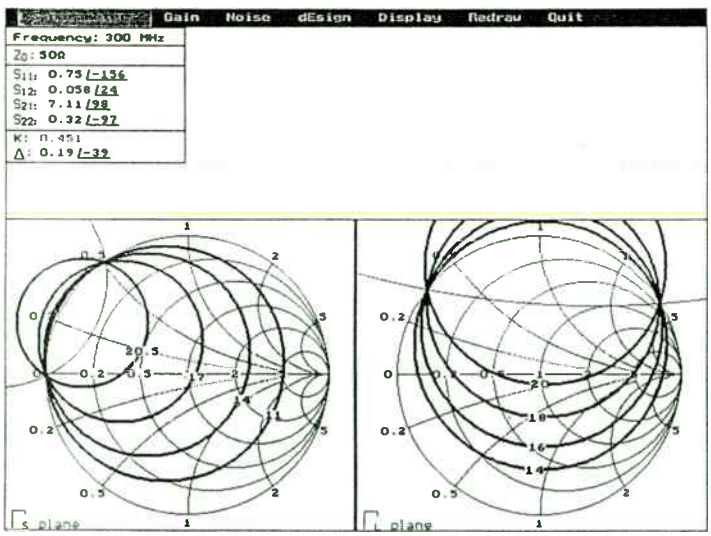


Figure 9. Constant gain circles

[3] G.D. Vendelin, *Design of Amplifiers and Oscillators by the S-parameter Method*, App. A, Wiley-Interscience, NY, 1982.

[4] D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in terms of S-parameters," *IEEE Transactions on Circuits and Systems*, February 1976.

[5] M. Vai and S. Prasad, "Computer-Aided Microwave Circuit Analysis by a Computerized Numerical Smith Chart", *IEEE Microwave and Guided Wave Letters*, vol. 2, No. 7, July 1992.

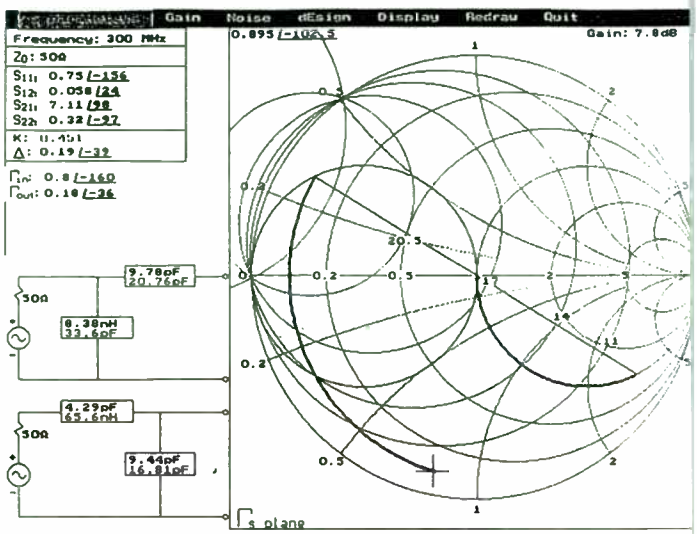


Figure 10. Input matching network menu

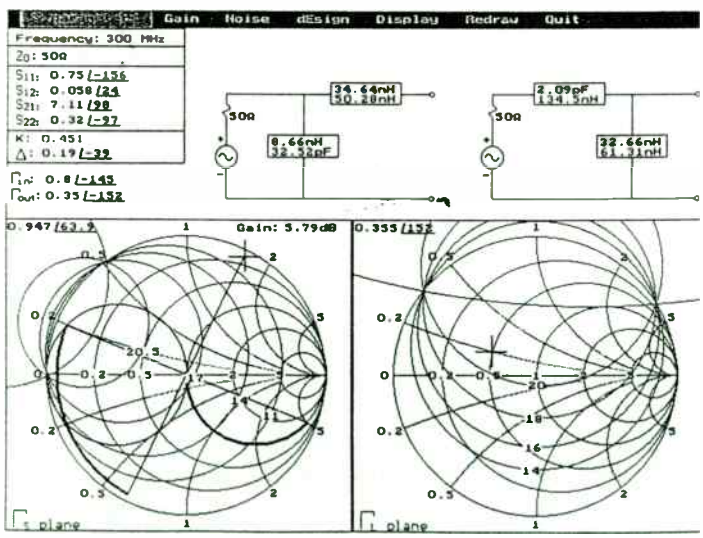


Figure 11. Input/Output matching network menu

SAW RESONATOR OSCILLATOR DESIGN USING LINEAR RF SIMULATION

Alan R. Northam
RF Monolithics, Inc.
4441 Sigma Road
Dallas, Texas 75244

The RF design engineer faces two basic problems when designing a Surface Acoustic Wave (SAW) Oscillator. The first problem the designer faces is how to properly model the SAW oscillator utilizing Computer Aided Design (CAD) software. The second problem is how to model and connect the SAW resonator in an oscillator circuit. This paper discusses the four ways the SAW resonator can be connected in an oscillator, and which configuration is the most appropriate to use. This paper also demonstrates that linear RF simulation is a very close approximation to the actual oscillator performance.

There are many design topologies in which oscillators can be configured. Among the most popular topologies are the Pierce, Colpitts, and Clapp. Each of these topologies have their advantages and disadvantages. It is not the intent of this paper to go into a discussion of each of these oscillator topologies. The purpose of this paper is to illustrate the four ways of configuring a two-port SAW resonator, to present the electrical differences between them, and then to determine the proper SAW configuration to be utilized given a certain oscillator type.

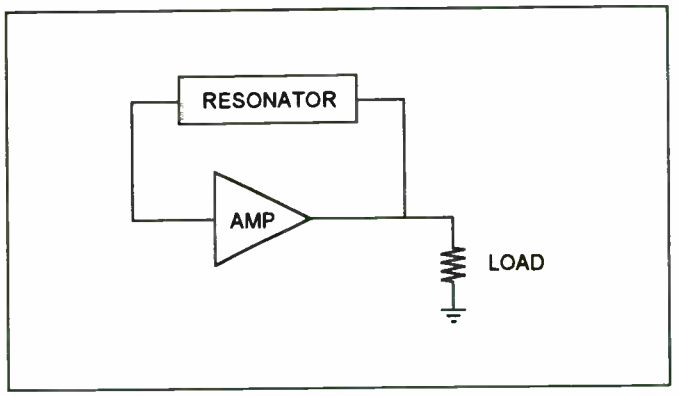


Figure 1. Pierce Oscillator

BASIC OSCILLATOR CIRCUIT

A simplified block diagram of a Pierce Oscillator is illustrated in figure 1. The basic oscillator loop consists of three basic elements: 1) An amplifier, 2) a feedback path and 3.) a load. The amplifier can be either of a discrete design, such as a transistor or FET, or it can be of a hybrid design, such as an Monolithic Microwave Integrated Circuit (MMIC). The feedback path typically contains a frequency selective network such as an LC resonant circuit, a crystal or, in the case of a SAW oscillator, a surface acoustic wave resonator. The load can either be a resistor or a complex impedance, such as the input of a following amplifier stage.

The basic criteria for oscillation in an oscillator are that: 1) the open loop gain must be greater than the losses around the oscillator loop and 2) the phase shift around the oscillator loop must be either 0 or 360 degrees.

In a typical oscillator, the amplifier starts off in its linear gain region, but as oscillation builds up the amplifier goes into its non-linear region as it goes into compression. The amount by which the amplifier goes into compression is a function of the loss in the oscillator feedback loop. As a minimum, the loss in the feedback loop is controlled by the loss of the SAW resonator, and the loss due to splitting the RF signal coming out of the amplfier into two paths, one path being the feedback loop, and the other path going to the Load. Additional loss can be established utilizing resistive pads, mismatch loss or lower gain transistors or MMIC amplifiers.

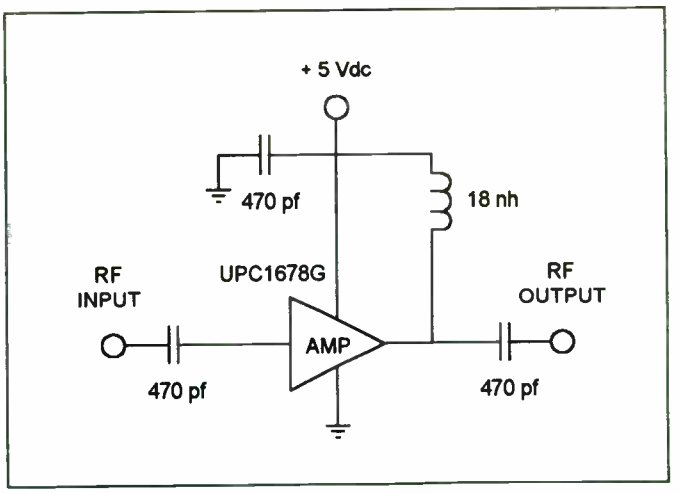


Figure 2. Amplifier Circuit

THE AMPLIFIER

The oscillator design example, as shown in figure 12, utilizes a MMIC amplifier. The amplifier circuit is shown in figure 2. The amplifier uses an NEC UPC1678G MMIC which has a fixed linear gain of 23 dB. The amplifier has a flat gain response with a 3 dB roll-off point of 1.9 GHz. The power supply voltage (Vcc) equals 5 volts dc and the bias current is 50 milliamps. The coil is used as an RF choke as well as a

path to provide high dc current to the output stage of the MMIC amplifier. This coil could also be used to match the output of the MMIC to a real impedance if desired. The input and output capacitors are used to block dc. The capacitor connected between the coil and ground is used to return the RF currents in the coil to ground.

Figure 3 is a graph of the MMIC amplifier gain. The theoretical transmission gain curve was plotted utilizing the scattering parameters generated for the amplifier design utilizing a linear RF simulator. The MMIC amplifier scattering parameters used in the simulation were those typically supplied by the manufacturer. The second curve is measured amplifier gain. Notice that the measured gain curve is virtually the same as the calculated gain curve and is approximately 17 dB. The third curve is measured gain with the amplifier 6 dB into compression. It is in this compressed condition that the amplifier will operate when the oscillator loop reaches equilibrium. The compressed gain of the amplifier is approximately 17 dB (23 dB - 6 dB = 17 dB).

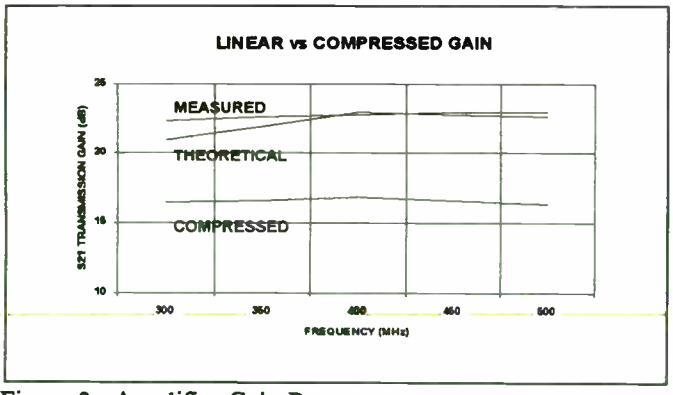


Figure 3. Amplifier Gain Response

In looking at the graph in figure 3, it is apparent that linear simulation of the amplifier is all that is needed to analyze the gain requirement of the oscillator. The designer can then easily determine the linear gain, using published scattering parameters and then offset the calculated gain by 6 dB to arrive at the compressed power gain. Using this approach to determining compressed gain can however lead to some error and the designer must determine if he or she wishes to accept that possibility. Notice, for example, that the measured linear gain and the theoretical gain deviate by about 2 dB at 300 MHz. Calculating compressed gain from the theoretical gain curve would lead to a 2 dB error or 15 dB (21 dB - 6 dB = 15 dB) versus the actual 17 dB (23 dB - 6 dB = 17 dB). The compressed gain curve in figure 3 was not generated by making such an assumption, it was generated using measured scattering parameters while maintaining the amplifier 6 dB in compression over a wide frequency range. Once these scattering parameters are measured they can then be used to design amplifiers over the entire frequency range characterized. When making these measurements it is desirable to also characterize the amplifier over a range of differing amounts of compression.

In looking at the amplifier phase plots in figure 4 notice that the difference in the calculated (theoretical) phase obtained from the published scattering parameters and the phase of the amplifier in compression is quite large: 50 degrees at the low frequency end and 30 degrees at the high frequency end of the graph. It is not possible to assume this phase offset by observation like it was in the case of the compressed amplifier gain of the MMIC. To determine the phase response of an amplifier in compression it becomes necessary to measure its phase response, or to utilize non-linear RF simulation software and an accurate model of the active device. Non-linear CAD software is expensive and in some cases is unnecessary in the design of RF circuits. Simple RF amplifier circuits involving one non-linear active device is such a case.

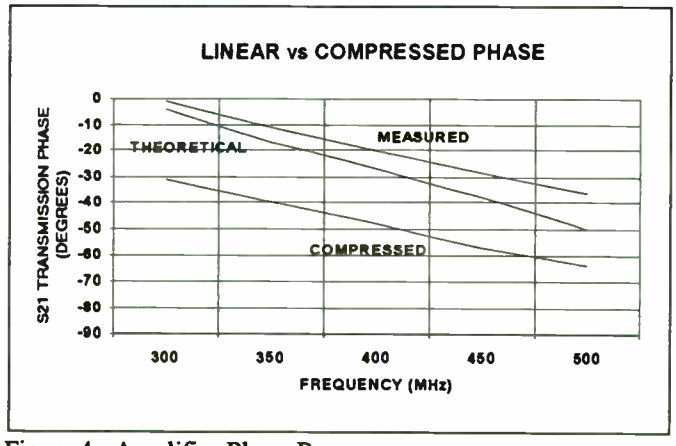


Figure 4. Amplifier Phase Response

While making phase measurements of the amplifier in compression, the designer should go ahead and measure all four scattering parameters. Once the compressed amplifier scattering parameters are measured it is then possible to determine not only the transmission gain and phase of the amplifier but also its input and output impedance. The compressed amplifier gain, phase, and matched input and output impedance, are the conditions under which the oscillator will exhibit proper loop gain and loop phase.

It is also under these conditions that the SAW resonator will be properly matched so as not to distort its passband characteristics. The amplifier's compressed input and output impedance is plotted on the graphs of figure 5 and 6 respectively.

The key to utilizing linear RF simulation involving non-linear devices is to characterize them under the conditions they will be used in the circuit. In the case of a transistor or MMIC used in an oscillator, the designer must measure its scattering parameters over, 1) the frequency range of interest, and 2) over a range of differing compression levels. These scattering parameters can then be used to design the amplifier stage used in an oscillator. The rest of the oscillator loop circuitry is easily designed using linear RF simulation as it is typically composed of linear network elements.

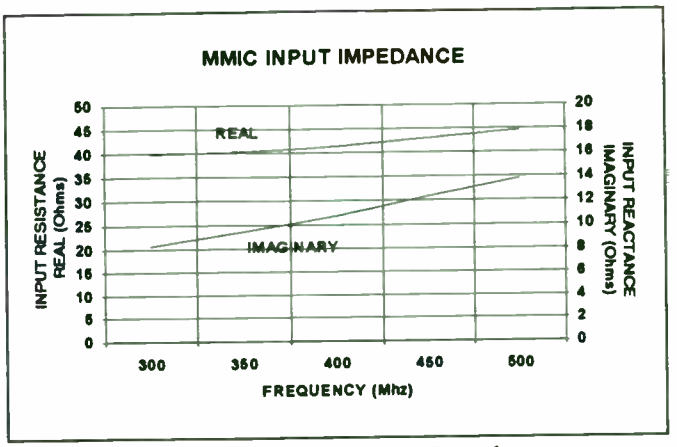


Figure 5. Compressed Amplifier Input Impedance

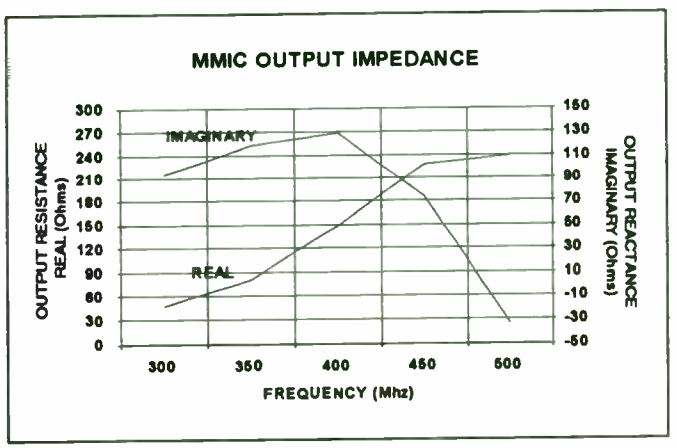


Figure 6 Compressed Amplifier Output Impedance

THE TWO-PORT SAW RESONATOR

The feedback path of the oscillator loop contains a frequency selective resonator network as shown in figure 1. In the case of a SAW oscillator this resonator is a SAW resonator. The SAW resonator is equivalent to a series LC network with a very high Q: a Q of 5000 or more is typical. The electrical model of a typical two-port SAW resonator is as shown in section (a) of figure 7.

The two-port SAW resonator can be connected in four different configurations as shown in figure 7. Section (a) of figure 7 shows the two-port SAW resonator connected such that its transmission phase is typically 180 degrees at its resonate frequency. The same two-port SAW resonator can also be connected such that its transmission phase at resonance is zero degrees. This is shown in section (b) of figure 7. The only difference between the SAW resonator having a transmission phase of zero degrees or 180 degrees is the selection of which output port terminal is grounded. From the electrical equivalent circuit shown in section (a) and (b), the proper transformer secondary winding is selected to set the transmission phase of the RF signal. The series LCR network represents the emotional inductance, capacitance, and the loss of the SAW resonator. The shunt capacitors represent the interdigital capacitance of the SAW structure. It is this capacitance that makes the SAW resonator input and output impedance capacitive even though the series LC network is at resonance. It is also this capacitance that keeps the resonators transmission phase from being either zero degrees or 180 degrees. These interdigital capacitances are quite often tuned out by resonating them with inductors to return the SAW resonator transmission phase back to either zero or 180 degrees. Tuning out the interdigital capacitance

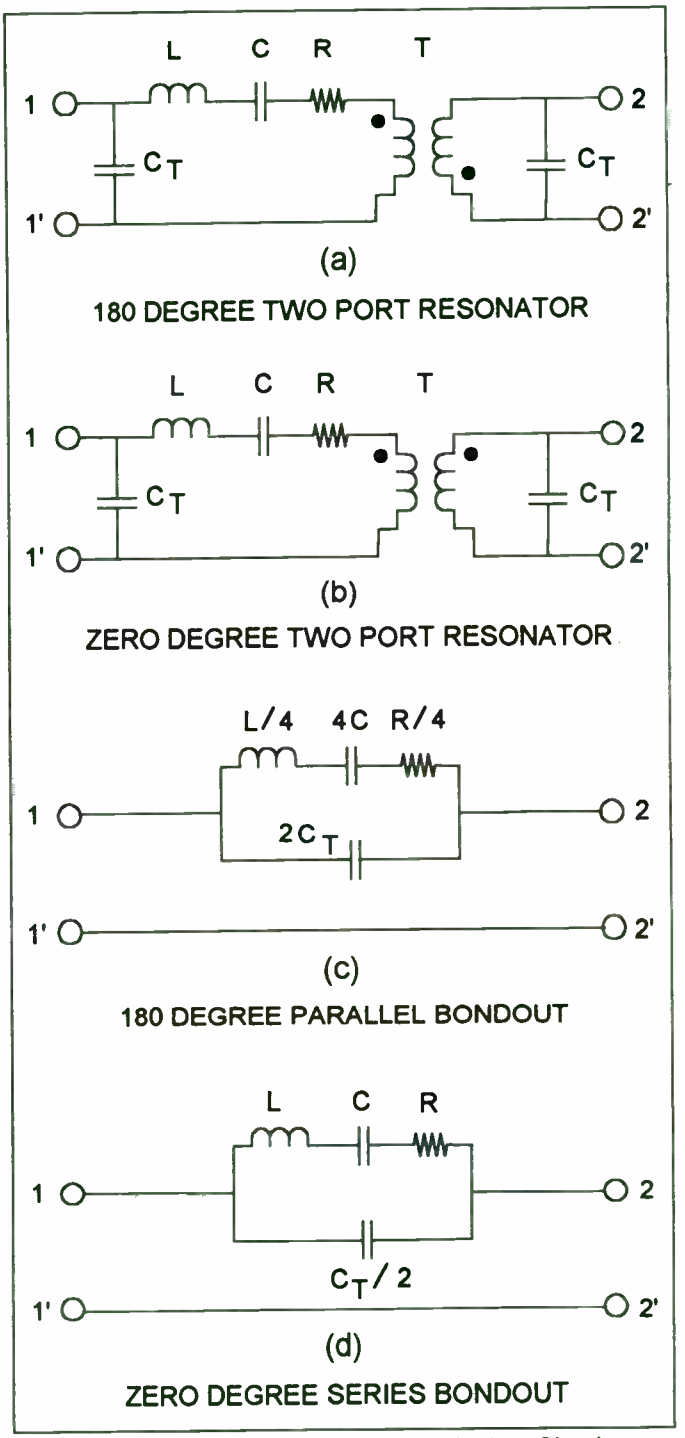


Figure 7. Two-Port SAW Resonator Equivalent Circuit

using series inductors will have the effect of reducing loss through the resonator at the expense of increasing its bandwidth, thereby decreasing its Q. Tuning out the interdigital capacitance using shunt inductors will not

degrade the loss through the resonator nor degrade bandwidth or Q. This results in the network impedance being real at the resonant frequency and is typically in the 50 ohm region for the series tuned inductor configuration, and in the 300 ohm region for the shunt inductor configuration.

The two-port SAW resonator can also be connected as shown in section (c) of figure 7. This configuration is created by taking the 180 degree SAW resonator connected as shown in section (a) of figure 7 and folding it in half, connecting terminal 1 to terminal 2 and terminal 1' to terminal 2'. The connection of terminal 1 to terminal 2 becomes terminal 1 of section (c), and the connection of terminal 1' to terminal 2' becomes terminal 2 of section (c). The return path, terminal 1' and terminal 2' of section (c) was added. This then connects the two interdigial capacitances C_T in parallel. Note in section (c) that the interdigital capacitance C_T is now twice the value or $2C_T$. Also notice that due to the transformer action the emotional inductance and the resistive loss is divided by four, and the emotional capacitance is multiplied by four.

The two-port SAW resonator can also be connected as shown in section (d) of figure 7. This configuration is created by taking the zero degree SAW resonator connected as shown in section (b) of figure 7 and connecting the input port in series with the output port. This is accomplished by connecting terminal 1' to terminal 2. Terminal 1 of section (b) remains unchanged and becomes also terminal 1 of section (d). Terminal 2' of section (b) now becomes terminal 2 of section (d). The connection of terminal 1' to terminal 2 of section (b) is an interstage connection in section (d) and is not shown in section (d). The return path, terminal 1' and terminal 2' of section (c) was added. This then connects the two interdigial capacitances C_T in series. Note in section (d) that the interdigital capacitance C_T is now half the value or $C_T/2$. Note that, due to the transformer action, the emotional inductance, capacitance and the resistive loss are not effected.

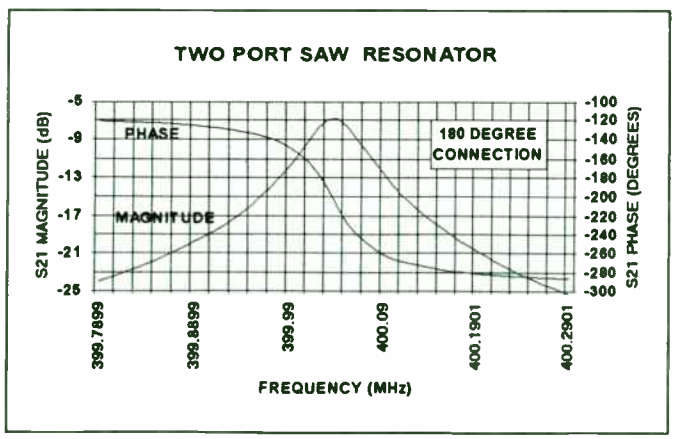


Figure 8. Two-Port Saw Resonator

The frequency and phase response of the two-port resonator connected as per section (a) of figure 7 is shown in the graph in figure 8. This graph was plotted from measured scattering parameters. Notice that the 3 dB bandwidth is approximately 60 kHz and the center frequency is 400.03 MHz. Calculating Q we find it to equal 6667 (400.03 MHz / 60 KHz = 6667). Also notice that the phase is approximately -200 degrees. Resonating out the interdigital capacitances with an inductor connected in parallel on both the input and output of the SAW resonator would bring the phase back to -180 degrees. (Series inductors will decrease Q and broaden bandwidth.) The SAW resonator has approximately 7 dB of loss and the phase change across the 3 dB bandwidth is approximately 90 degrees.

The input and output impedance's of the two-port resonator connected as per section (a) of figure 7 are shown in the graph in figure 9. This graph was plotted from measured scattering parameters. Notice that at the resonant frequency of the SAW resonator the reactance is at its minimum value and increases rapidly after passing through resonance. Also notice that the resistance of the resonator is close to 100 ohms and also increases rapidly after passing through resonance. By tuning out the interdigital capacitance of the SAW resonator with an inductor will reduce the input and output reactance to zero ohms, leaving the input and output impedance under this tuned condition real and approximately 100 ohms if series inductors are used and 300 ohms if shunt inductors are used.

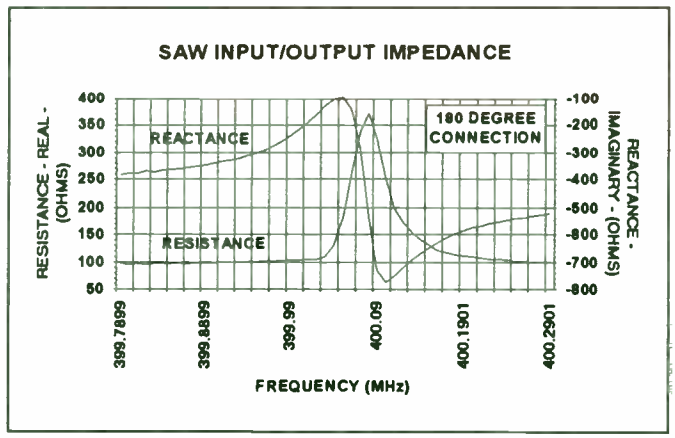


Figure 9. Two-Port Resonator Input / Output Impedance

The two-port SAW resonator connected as shown in section (b) and (d) of figure 7 yields virtually the same magnitude and phase response as shown in figure 8 and 9 and will not be discussed. The only difference is that the transmission phase is close to zero degrees at resonance for both the (b) and (d) connections provided the interdigital capacitances are tuned out. Notice that only a single inductor connected from terminal 1 to terminal 2 in section (d) of figure 7 is needed to tune out the interdigital capacitance. Tuning out the inter-digital capacitance of section (d) does not effect bandwidth of loss.

When the same two-port SAW resonator is connected, as shown in section (c) of figure 7, the resonator exhibits lower loss and broader bandwidth. The lower loss is expected because its loss being divided by four, R/4. However the increased bandwidth may not have been expected by the

casual observer. Increased bandwidth occurs because the 50 ohm source and load impedance environment remained constant while the resistive loss R of the resonator was reduced by a factor of four. Had the 50 ohm source and load environment changed proportionally the bandwidth of the resonator would have remained unchanged.

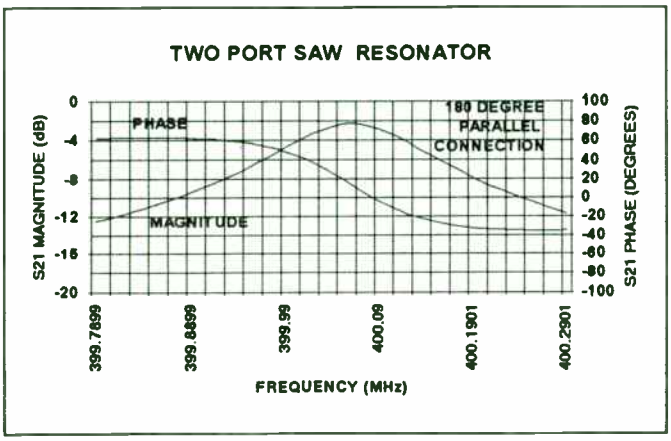


Figure 10. Two-Port SAW Resonator (Section C of Figure 7)

Figure 10 is a plot of the magnitude and phase of the two-port resonator connected as shown in section (c) of figure 7. Notice that the loss has dropped to 2 dB and the bandwidth has increased to 170 kHz. Calculating Q yields a Q of 2353. Notice that the phase slope has become more shallow while still exhibiting approximately 90 degrees across the 3 dB bandwidth.

Decreased Q equates to less stability in an oscillator. Although a Q of 2353 is still quite high and the oscillator will still appear to be quite stable, in a high stability oscillator design connection (c) of figure 7 would not be the best choice.

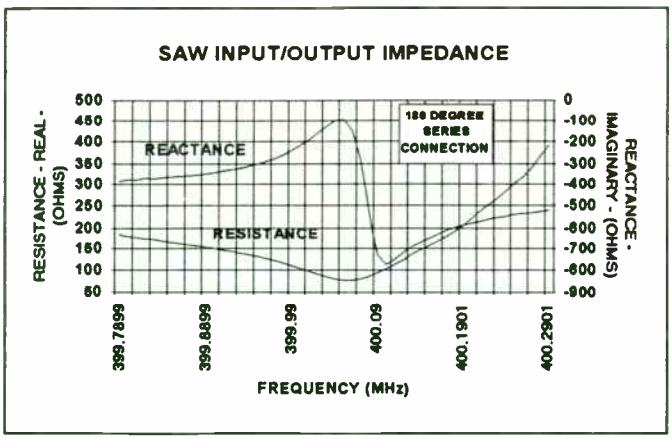


Figure 11. Two-Port SAW Resonator (Section C of Figure 7)

The input and output impedance for the Two-Port SAW Resonator connected per section (c) of figure 7 is plotted in the graph in figure 11. Notice that at resonance the real resistance is closer to 50 ohms than any of the other three connections, hence its lower loss. Notice also that the capacitive reactance curve has a similar shape to the other three connections and is again at a minimum at resonance.

From the four different ways of connecting a two port resonator, three of the configurations yield the same magnitude response, section (a), section (b), and section (d). A SAW resonator connected as in section (c) of figure 7 offers lower loss but at the cost of increased bandwidth. A SAW resonator connected as shown in section (b), section (c), and section (d) of figure 7, has a transmission phase of zero degrees at the resonant frequency and exhibits 90 degrees of phase shift across its 3 dB bandwidth. The SAW resonator connected as per section (c) exhibits a more shallow phase slope than the other two connections due to its wider bandwidth. A SAW resonator connected as shown in section (a) is the only configuration that offers -180 degrees of transmission phase at resonance.

Deciding on which SAW resonator configuration to use comes down to basically two choices: 1) A SAW resonator connected as per section (a), if the phase shift through the SAW resonator need be -180 degrees to satisfy the oscillator loop phase requirement, and 2) The SAW resonator connected as per section (d), if the phase shift through the SAW resonator need be zero degrees to satisfy loop phase requirements. The SAW resonator connected as per section (b) could be used instead of section (d) but would require two inductors to resonate out the interdigital capacitances, whereas connected as per section (d) requires only one inductor. From a printed circuit board real estate and a cost viewpoint the connection shown in section (d) of figure 7 is most likely the better choice. The configuration of section (c) would be a good choice if lower loss were required, or broader bandwidth to speed up turn-on time. Another application of the configuration of section (c) would be for voltage controlled oscillators where broader frequency tuning is required.

OSCILLATOR DESIGN EXAMPLE

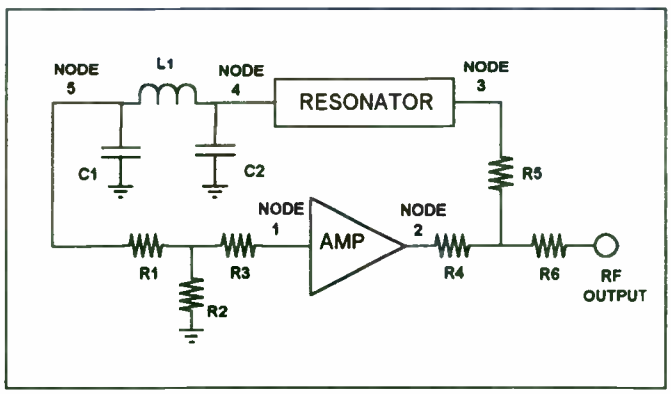


Figure 12. 400 MHz Fixed Tuned Oscillator

The oscillator design example is shown in block diagram form in figure 12. The amplifier circuit is shown in further detail in figure 2. A 6 dB resistive power splitter made up of R4, R5, and R6 is connected to the amplifier output. One output of this divider is connected to the RF output while the other output is connected to the feedback path. The resonator is a two-port SAW resonator configured for the proper

transmission phase. Notice that in this design example the SAW resonator's interdigital capacitance is not tuned out by resonating them with parallel or series inductors. The LC pi-network is used to set the loop phase to zero degrees. The resistive T-network is used to set the loop gain to 6 dB. The LC pi-network and or the resistive T-network may or may not be necessary depending upon frequency, the amplifier used, or how accurately the designer wishes to set loop gain and phase.

In low cost circuits the resistive pad used to set loop gain precisely and the phase setting LC network are compromised or left out completely. This decision may or may not cause problems to the performance of the oscillator. An oscillator designed without regard to setting loop phase properly will result in an oscillator running off the center frequency of the SAW resonator. Depending upon the amount of phase error, the oscillator loop gain can be reduced due to the oscillator running too far off the center frequency of the SAW resonator and the resonator acting like the bandpass filter that it really is. If loop gain is not set properly then the oscillator could have too much loop gain that it will oscillate off frequency due to the SAW resonator's other modes. In addition, to much loop gain can cause the amplifier to be overdriven, causing the output power to foldback.

The oscillator loop was modeled using a linear RF simulator. The compressed amplifier scattering parameters were input to an ASCII file of the proper format for the RF linear simulator used. The resistive power splitter was then modeled and cascaded to the amplifier output at node 2 in figure 12. A two-port network consisting of scattering parameters of a SAW resonator was also modeled and cascaded to the power splitter at node 3. A linear RF simulation analysis was then run from node 1 to node 4 to determine the overall gain and phase of the oscillator loop thus far. The result showed that the loop still had an excess of 5.5 dB of gain and that the phase was -261 degrees. The phase setting pi-network consisting of C1, L1 and C2, and the resistive T-pad network consisting of R1, R2 and R3 were then modeled and cascaded to the output of the SAW resonator node 4 of the open loop model of the oscillator loop. The linear RF simulator was then programmed to optimize for a loop gain of zero to 1 dB and loop phase of zero to 1 degrees. After the optimization was complete the closest standard values for R1, R2, R3, C1, C2 and L1 were chosen and a final simulation analysis performed. The result of the final simulation analysis of the oscillator loop yielded a loop gain of -0.6 dB and loop phase of 4.3 degrees. Installing the final component values in the oscillator should cause the oscillator to operated in the middle of its passband of the SAW resonator at 400.03 MHz (refer figure 8), and the amplifier should be 6 dB in compression.

The oscillator output power cannot be predicted using a linear RF simulator, so it becomes necessary to employ a different technique to make this prediction. The output power of the amplifier (refer to figure 2) is measured for different input power levels applied. This data is then input to a computer application spreadsheet program and graphed as shown in figure 13. From the graph the designer can determine the output power of the oscillator. Alternately, the data can simply be graphed using graph paper. The amplifier output power should be characterized over several different frequencies across the frequency spectrum of interest. Once this full matrix of graphs is created the designer can interpolate between the frequencies graphed to determine output power at a specific frequency. With the computer applications spreadsheet program, the designer can input equations into the spreadsheet to do the interpolation. Once the spreadsheet is set-up, and all the equations entered, it can be used as a template for other amplifier circuits the designer wishes to characterized.

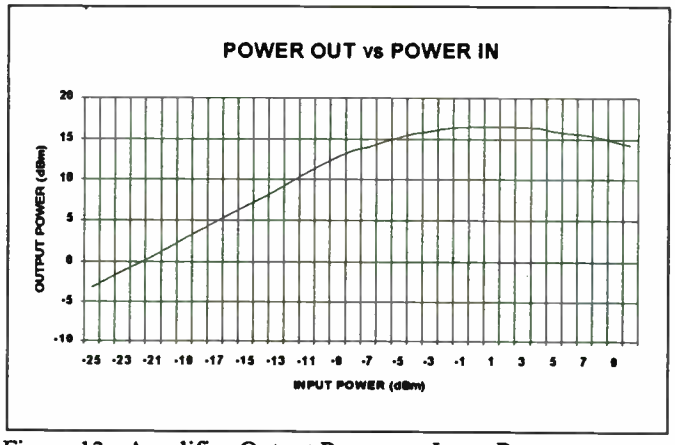


Figure 13. Amplifier Output Power vs. Input Power

From the graph of figure 13 it can be seen that the amplifier output power is at its maximum when the input power is somewhere between -3 dBm and +3 dBm, and is approximately 17 dBm. From the previous discussion of the oscillator design example, it was determined that the oscillator loop was to have a loss of 17 dB. Since the amplifier output compresses at 16 dBm the power fed back to the amplifier input will therefore be -1 dBm (17 dBm -17 dB = 0 dBm). From the graph of figure 13, it is seen that with the input power of the amplifier at 0 dBm, the output power of the amplifier will be close to its peak at 17 dBm.

Notice that too much power applied to the input of the amplifier causes the output power to foldback thus decreasing more and more with increased input power. This is a typical phenomena of an RF amplifier being driven too hard. This phenomena is also a cause of error in predicting the output power of an amplifier when loop gain is improperly set as previously discussed.

OSCILLATOR PERFORMANCE

The oscillator design example was built using component values determined with the linear RF simulator optimizer. The oscillator's actual frequency of oscillation was 400.022 MHz, or only a 8 kHz off the center of the SAW resonator's

400.03 MHz center frequency. This error is well within expectations and equates to a phase error of only 11 degrees.

phase error = frequency error * delta phase shift / bandwidth

11 degrees = 8 kHz * 90 degrees/ 65 kHz

The delta phase shift term in the phase error equation refers to the change in transmission phase across the 3 dB bandwidth of the SAW resonator.

Recall that in selecting standard values, after optimizing the oscillator loop using the linear RF simulator, the final analysis showed that there would be a 4.3 degree phase error. Thus the 11 degrees of phase error determined by calculation due to the actual closed loop oscillator frequency measurement is actually only a 7 degrees from what the linear RF simulator predicted.

The oscillator output power is calculated as follows:

P_o = Max Amplifier P_o (dBm) - Power Splitter loss (dB)

11 dBm = 17 dBm - 6 dB

The oscillator performance is shown in figure 14. Notice that the output power measures to be 12 dBm and is only 1 dB different than the calculation. Some of this error is due to the accuracy of interpreting the actual amplifier output power from the graph in figure 13.

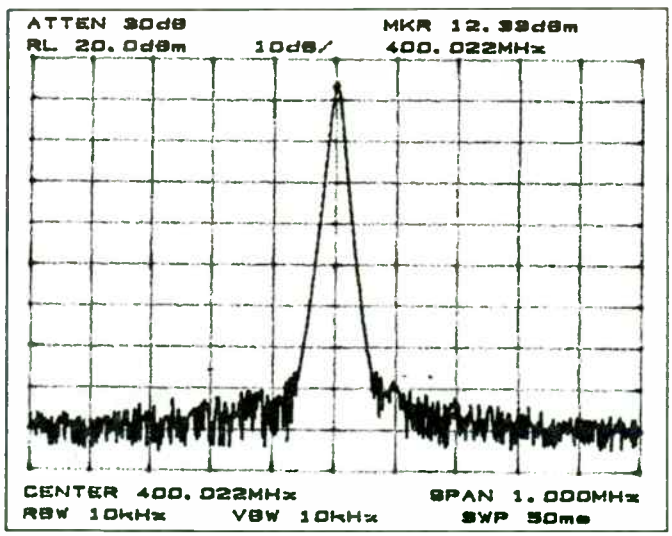


Figure 14. Actual Oscillator Performance

CONCLUSIONS

There are four different ways of connecting a SAW resonator as a two-port device. Each configuration has its advantages and disadvantages. Two of the configurations require two inductors to resonate or tune out the interdigital capacitances, while two of the configurations only require one inductor. One of the four configurations have a transmission phase of -180 degrees while the other three has an S21 transmission phase of zero degrees. One of the configurations has reduced transmission loss but at the expense of increased bandwidth.

The SAW resonator interdigital capacitance can be tuned out by using series inductors of parallel inductors. Series inductors have the effect of reducing loss but causes bandwidth to increase, while parallel inductors do not reduce loss nor increase bandwidth.

The key to using linear RF simulation is to characterize the amplifier while in its compressed state. The designer should measure all four scattering parameters over the frequency range of interest. These scattering parameters are then used with the linear RF simulator to predict loop gain, loop phase, and input and output impedance.

Open loop linear RF simulation of the oscillator, utilizing compressed amplifier scattering parameters, accurately predicted the oscillating frequency of the oscillator. It also calculated the component values needed to meet specific performance goals, i.e. loop gain and loop phase.

Output power can also be predicted quite accurately by measuring the output power of the amplifier at various input power levels and then graphing the result. Graphs of output power versus input power should be created for various frequencies across the frequency spectrum of interest. This data can then be used to interpolate the output power expected at a particular frequency. A similar technique can also be used to predict harmonic levels.

The linear RF simulator is an appropriate tool for modeling non-linear oscillator circuits. As shown in the design example, quite accurate results can be achieved in predicting the oscillator frequency and output power. The key to its accuracy is in characterizing the amplifier over power input levels and over frequency. Once this characterization is complete the data can be used to design oscillators over a wide frequency range.

ACKNOWLEDGEMENTS

I gratefully acknowledge the efforts of Lonnie Harmon for characterizing many two-port SAW resonators and building and testing the prototype oscillator, and Christina McFarland for editing the manuscript and coordinating all the efforts necessary to make this paper possible.

REFERENCES

1. Robert J. Kansy, "Introducing the Quartz Surface Acoustic Wave Resonator," 1983 RF Monolithics Application Note No. 1 page 13.

2. Marvin E. Frerking, "Crystal Oscillator Design and Temperature Compensation," Van Nostrand Reinhold Company Inc, New York, c. 1978, page 58

Embedding RF Design Tools in an IC Design System

Mojy C. Chian, Steve S. Majors, and Alan G. Whittaker

Harris Semiconductor
Melbourne, FL 32901

Abstract *-The recent introduction of IC technologies offering high frequency transistors with f_t in the vicinity of 10 GHz has opened new opportunities for higher integration of wireless communication systems. Fast silicon IC devices make possible the integration of many RF subsystems on a single die and offer a total solution to mixed frequency (low frequency and RF) and mixed signal systems. To realize this opportunity, IC design systems have to be enhanced to accommodate RF specific design tools. This paper presents a unified design environment for combined IC and RF designs. The approach taken was to enhance and modify an IC design framework, simulation engine, and data analysis tools to provide RF specific design and data representation capabilities.*

I. Introduction

This decade will foster a growing and rapidly changing market segment in wireless communication products and lightwave components. The driving force behind the fast pace is the search for lower cost, smaller size, and higher performance products. However, traditional discrete based designs are unable to keep up with the pace. Discrete based designs are quickly reaching the physical limits on size, package and interconnect parasitics, and electrical performance.

RF designs are typically comprised of many individual subsystems such as low noise amplifiers, mixers, filters, and automatic gain controls. Traditionally, RF system designers have realized the individual subsystems with discrete components. With the emergence of IC processes offering transistors with an f_t around 10 GHz, silicon ASICs are offering a higher level of integration of RF designs. A system level solution, in which many RF subsystems are integrated on a single silicon die promises dramatically smaller size and in many cases higher performance.

The other major advantages of silicon ASICs over discrete designs are customizable transistors and predesigned cell libraries. RF and low frequency microwave circuits typically require customized transistors to optimize gain, minimize noise figure, and reduce distortion. Predesigned cells in an ASIC vendor's cell library can drastically reduce the overall design time of an RF system. IC foundries offering high frequency processes, variable geometry devices, predesigned cells, and a front-to-back design system will prove to be the essential ingredient for reducing the cost, size, and time-to-market for RF systems.

To realize opportunities in the new and expanding wireless market, an efficient methodology for the design of RF ASICs is needed. Reliable and accurate tools for predicting the ASIC performance before fabrication is essential, since breadboarding is not possible and refabrication is very costly. Historically, IC and RF designers have used different design goals, design methodologies, and practices [1-3]. As the boundary between IC and RF design blurs, both IC and RF designers are compelled to design in each others domain. Designers in both areas are beginning to recognize the need for a CAD system that supports the design tools of both domains. The combined system should provide a consistent design environment in which both IC and RF designers will find their familiar tools and user interface as well as the capability to easily traverse to the other environment.

This paper describes FIRST (Fastrack's Integrated RF Simulation Tools). FIRST is a set of embedded RF design tools in an IC design system aimed at the design requirements for RF designs in silicon ASICs. The following sections elaborate on the structure and capabilities of FIRST, and demonstrate its use on a production circuit.

II. System Architecture

There are two possible methodologies to combine IC and RF design systems in one environment.

Direct Integration

With direct integration of IC and RF design tools, the IC and RF design systems are merged together to form a superset system. The direct integration of the two design

systems may seem to be an obvious solution. Unfortunately, because of little commonality between the two systems, direct integration has proven to be impractical. The two systems typically use different frameworks, simulation engines, and data analysis routines. To further complicate the problem, the cell libraries and device models are also different. The device models (which generally are not generic Spice models) must be ported, supported and maintained in two or more different simulators. Another factor is the price. Typically IC and RF design systems are owned by different vendors, which raises the cost of the combined system.

Embedded Tool Methodology

An alternate solution is to embed RF design tools in an IC design environment. In an embedded (or native) tool methodology the capabilities of one system are replicated inside the other system utilizing the same analysis engines. IC frameworks have typically matured much more than RF frameworks and thus can better serve as the host. In this case, the IC framework, simulation engine, data analysis tools, etc. are enhanced and modified to provide RF specific design tools. In an embedded methodology, the simulation data base and cell libraries are the same, regardless of the type of application (RF or IC). Also, a single analog simulator serves as the common simulation engine. This greatly reduces the device modeling problems. It is easier to implement a new model and easier to upgrade and maintain the existing device models. There is never an issue with model consistency since there is only one simulator.

The structure of FIRST

FIRST is a set of RF design tools integrated into FASTRACK [4], based on the embedded tool methodology. FASTRACK is a complete front-to-back (schematic-to-layout) IC design system (Figure 1). cdsSpice [5], an enhanced derivative of SPICE [6], serves as the simulation engine.

The raw data for RF results is generated from internally controlled single or multiple simulations. The netlist for each simulation is automatically generated according to the particular setup requirements of that analysis (see next section). The raw simulation data (e.g. S parameters) is saved in cdsSpice memory. The user can then choose the desired RF results (e.g. stability circles). A unified post processor uses the raw simulation data, performs the required mathematical processing to generate the user requested results and displays them on an appropriate plot format (e.g. Smith chart).

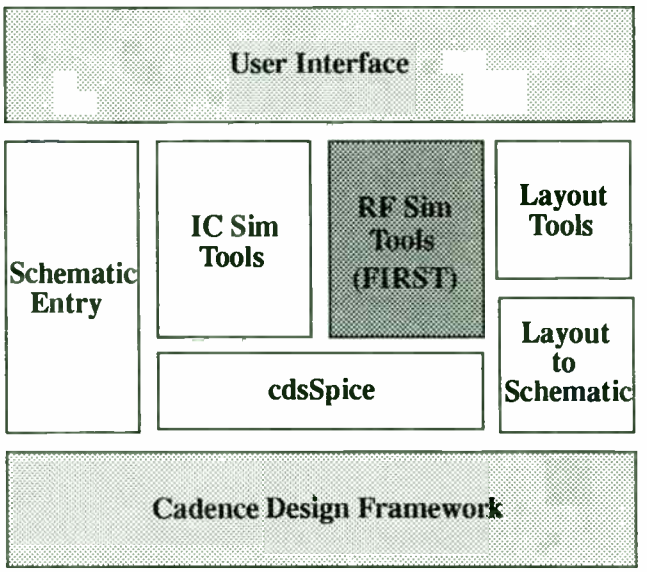


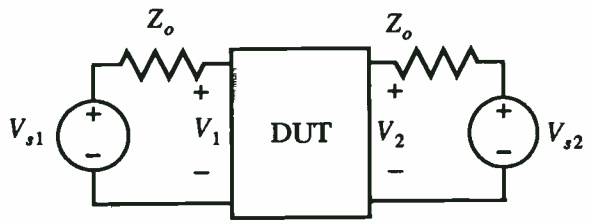
FIGURE 1. The structure of Fastrack, showing the embedded RF design tools, FIRST.

III. System Capabilities

Many RF specific design tools fundamentally use the same mathematical basis and numerical algorithms that exist in IC tools. The procedure to extract the required data can be different and it may require a special set up and a controlled environment. The user interface, analog simulation engine (cdsSpice), and the data representation routines in FASTRACK were enhanced and modified to implement FIRST. The major components of FIRST are described below.

S parameter measurement

S parameter measurement enables the designer to measure all four S parameters [7] for a single transistor or a two port network. The system establishes the proper loading and excitation (also auto biases single transistors) (Figure 2), runs multiple simulations and mathematically processes the results to determine the S parameters [8]. Once the S parameters are determined they can be printed to the screen or to a user specified file in any user specified format. They can also be plotted on Smith charts, polar, or rectangular plots. A full-function graphical environment allows the user to change plot axes, do multiple overlay plots, zoom into a specified region of the plots, etc. The S parameters are also used to generate constant power gain and stability circles, plot and print K factor, input and output VSWR, and input and output impedances.



Test #1

$V_{s1} = 1$

$V_{s2} = 0$

$S_{21} = 2V_2$

$S_{11} = 2V_1 - 1$

Test #2

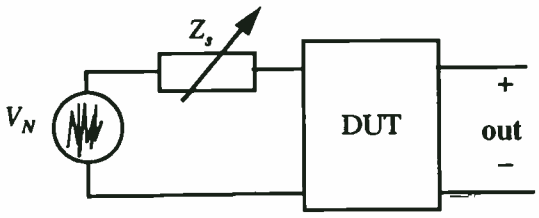
$V_{s1} = 0$

$V_{s2} = 1$

$S_{22} = 2V_2 - 1$

$S_{12} = 2V_1$

FIGURE 2. S Parameter measurement circuit.



$$F = 10\log\left(\frac{S_T}{S_s}\right)$$

$$S_T = \sum_{\text{all noise components}} S_i$$

FIGURE 3. S Parameter measurement circuit. Noise figure (F) is defined as the ratio of the total available noise power at the output (S_T) to the available noise power at the output due to the input noise (S_s).

S Parameter Modeling and Simulation

Frequently, RF designers require the capability to represent a two-port network by its S parameters in a tabular form. In the embedded RF tools, a generic macromodel provides the capability to represent the external behavior of any two-port network using an S parameter file. This file can either be generated by the S parameter measurement utility in FIRST or from a data sheet for other vendor's parts. The modeling is based on the equivalent two port Y parameters which in turn are generated from the complex tabular S parameters.

Noise parameter measurement

RF designs are frequently driven by noise specifications. Even though, IC and RF designers basically use the same noise models and AC small signal noise analysis, the interpretation of the results, the type of requested information, and the noise information processing are different. RF designers are typically interested in noise figure at a given frequency, minimum noise figure, noise resistance, and optimum reflection coefficients versus frequency. The noise parameter measurement capability in FIRST enables the designer to measure RF specific noise parameters for a single transistor or a two port network. As with S parameter measurement, the system automatically connects a bias circuit to a single transistor. The system places a noise source and an internally controlled impedance at the input of the circuit and runs 4 simulations to determine noise figure, minimum noise figure, noise resistance, and optimum reflection coefficient versus frequency [9].

Noise figure, minimum noise figure, and noise resistance versus frequency can be plotted or printed (to the screen or a user specified file). Optimum reflection coefficient can be printed or plotted on a rectangular plot or a Smith chart. The user can also plot the constant noise circles on a Smith chart. In addition, constant noise circles can be overlaid on the constant gain or stability circles to allow for a visual design trade off between gain, stability, and noise performance of a circuit.

Large Signal AC

This analysis enables the RF designer to observe the nonlinear performance of a circuit in the frequency domain. It can be used for single tone circuits to observe harmonic distortion or for multitone circuits to observe the intermodulation products. The third order intercept point is automatically calculated from the frequency spectrum. The power spectral density can be plotted (in a spectrum analyzer type bar plot) or printed (to the screen or to a file). The system determines the desired frequency spectrum by invoking an internally controlled and automated nonlinear transient analysis followed by a Fast Fourier Transform (FFT). The transient analysis and FFT are designed to internally control each other for optimum accuracy and efficiency (Figure 4).

While most time domain analysis and FFT based methods only offer about 60-80 dB of dynamic range, this new method can offer up to 260 dB of dynamic range. This methodology is totally different from Harmonic Balance, even though both methods use a combination of FFT and nonlinear transient analysis. Harmonic Balanced based simulators are more suitable for smaller circuits (less

than 10 transistors) that either take a long time to reach steady state (e.g. switching power supplies) or when the beat frequency in a multitone application is several (6 or higher) orders of magnitude smaller than the tones. While, Harmonic balanced based simulators become extremely inefficient for anything but very small circuits, large signal AC analysis in FIRST is as efficient as typical SPICE transient analysis for larger circuits.

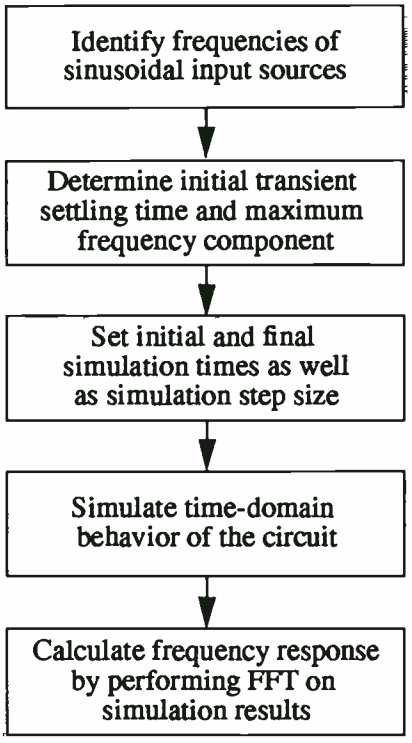


FIGURE 4. Flow of large signal AC analysis.

Small Signal Distortion

This analysis enables the designer to observe the nonlinearity and distortion for quasi-linear circuits. Quasi-linear circuits are a class of circuits that are ideally linear. In reality, they are almost linear but exhibit some undesired nonlinearities (e.g. an LNA). Small signal distortion analysis [10, 11] exploits this nonlinearity by determining the undesired effects of each nonlinear component in the circuit on an output load. This analysis uses linear transfer functions and, consequently, it is significantly faster than FFT based analyses. At a given fundamental frequency, the frequency spectrum, comprised of the major intermodulation components, can be plotted or printed. The fundamental frequency can also be varied over a specified range, and the intermodulation components can be plotted or printed versus the fundamental frequency. The third order intercept point can also be automatically calculated.

Backend Tools

The RF specific analysis tools and data representations in FIRST provide a familiar working environment throughout the circuit design stage. Equally important, the physical design and verification tools in FASTRACK provide a smooth transition from electrical design to physical design so that the success of the silicon ASIC is not compromised by the level of understanding of the IC technology. This system offers a comprehensive set of physical design automation tools that allow the user to create a full-custom layout, while operating at a high level of abstraction.

Parameterized device libraries are the basis for the layout strategy. These device layouts are designed to fully implement the variable geometry structures that were modeled during simulation; they are not limited to discrete values. In effect the parameterized cells completely remove the numerous IC device level design rules from the layout process, without loss of freedom or functionality. This is most easily appreciated by observing the cross-sectional area of a Harris UHF-1 transistor (Figure 5). Each interface between process steps is the source of a design rule that is critical to the success of the process flow.

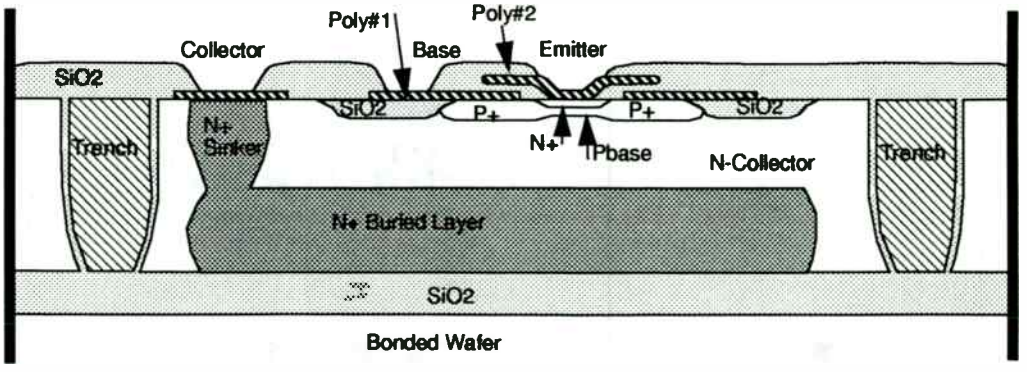


FIGURE 5. Cross-section of Harris UHFN1 transistor showing numerous process interfaces that must conform to design rules

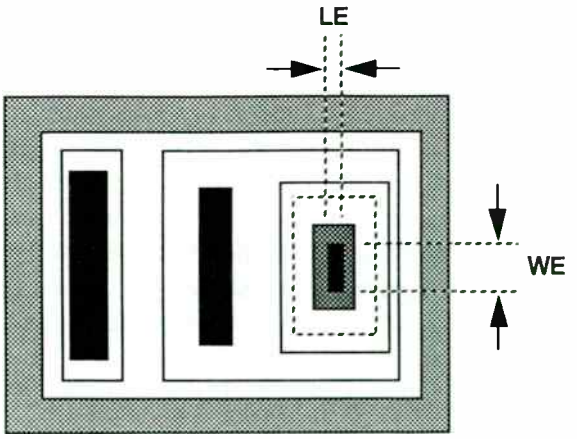


FIGURE 6. Simplified top view of Harris UHFN1 transistor showing emitter length (LE) and emitter width (WE) electrical parameters

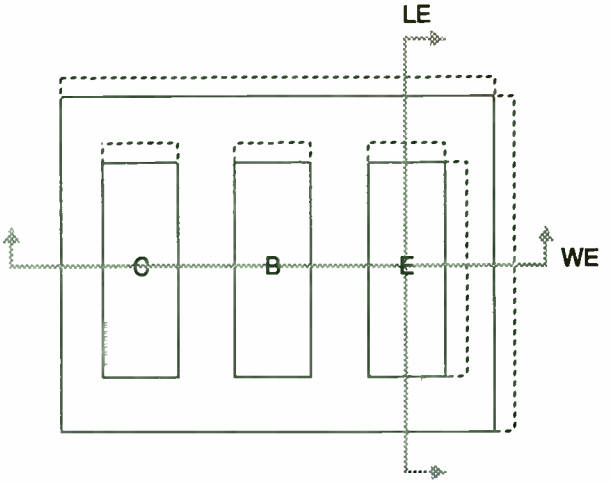


FIGURE 7. Effective parameterized cell of Harris UHFN1 transistor showing stretch lines that adjust internal geometries based on electrical parameters

The device layouts are automatically synthesized from the schematic by placing the parameterized cells relative to their schematic positions and applying the corresponding model parameters to size the geometries. In this way, the transistor in Figure 6 can be thought of as a virtual black box, as shown in Figure 7, with terminals for connecting the collector (C), base (B), and emitter (E). The device can now be thought of as having stretch lines that bisect the layout structure to adjust all of the internal geometries as a function of the electrical parameters (e.g. emitter length, emitter width, etc.). A wide range of layout optimizations are also supported by simply changing device parameters; for example, parallel and serpentine resistor structures, trimmable thinfilm (as a function of trim range and trim sensitivity), and capacitor aspect ratio. This allows the user to concentrate on just the circuit level interconnections.

Verification of the completed IC layout is accomplished with a Design Rule Checking (DRC) tool to ensure that the layout conforms to all manufacturing specifications, (metal width, metal spacing, etc.), and a Layout Versus

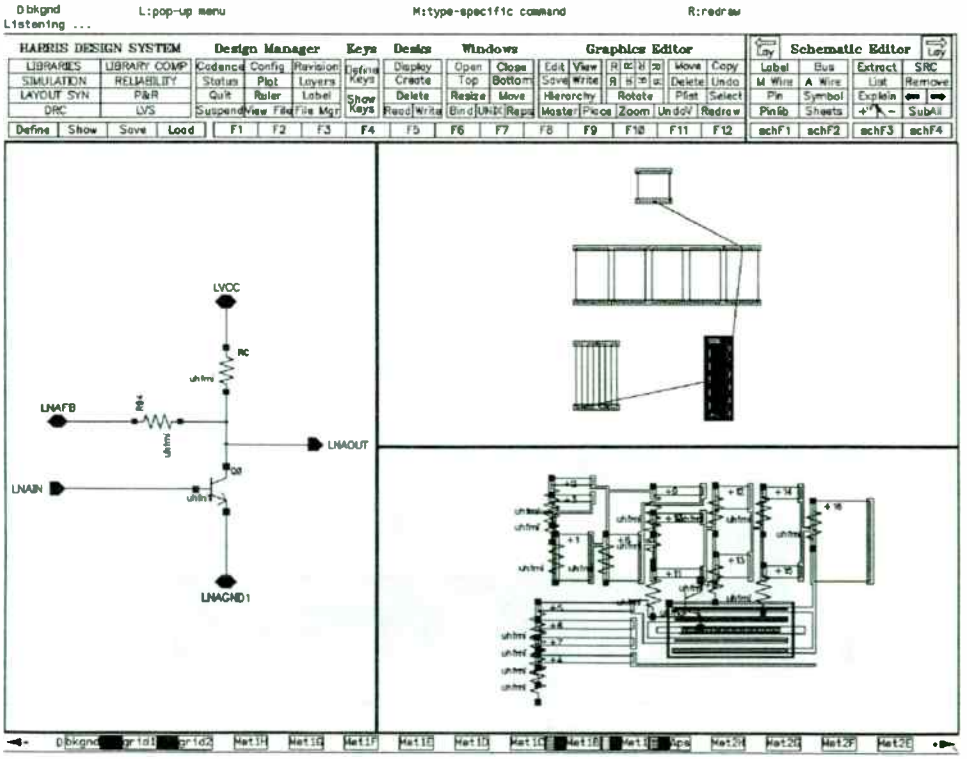


FIGURE 8. A screenshot of FASTRACK, showing the schematic, "flight-lines", and the extracted schematic overlaid on the layout.

Schematic (LVS) tool to validate the electrical functionality of the layout with respect to the schematic. DRC highlights portions of the layout (Figure 8, top) that fail to meet manufacturing guidelines and provides detailed information on how to correct the errors in terms of layout dimensions. LVS matches the connectivity between the interconnections of the devices in the schematic (Figure 8, left) and the electrically equivalent interconnections between the extracted devices in the layout (Figure 8, bottom). The layout extraction identifies the devices and the interconnections between devices by recognizing the diffused patterns and relational interactions among all of the different layout mask layers. The resulting extracted view of the layout (Figure 8, bottom) includes all of the electrical and physical attributes of the circuit in a form that can be compared with the schematic. This process requires no input from the user regarding the IC technology or the design application.

The results of the comparison between the layout and schematic describe differences in the interconnect paths (e.g. opens, shorts), the device types, and the device sizes. The layout and schematic networks are automatically reduced to minimum equivalent circuits so that, for example, a single resistor in the schematic can be represented by a complex series/parallel configuration in the layout and the two views will still match electrically. Throughout the layout and verification process, a tight correspondence is maintained between devices and interconnections in different representations. This allows "flight-lines" to be drawn in the layout showing what connections need to be made to implement the schematic design (Figure 8, top). It also provides cross-probing between all of the circuit representations after LVS.

Once the layout is verified, parasitic resistances and capacitances can be measured and back-annotated to the schematic netlist for simulation of the parasitic effects on circuit performance.

IV. Results

The low noise amplifier (LNA) shown in Figure 9 was used to demonstrate the analysis capabilities of FIRST. The LNA is part of a Harris LNA/Mixer IC designed to operate around 900 MHz.

The S parameter measurement capability of FIRST was used to determine the forward and reverse reflection parameters, S11 and S22 and plot them on a Smith chart (Figure 10).

The same analysis tool generated the data required to calculate the forward and reverse transmission parameters, S12 and S21. A polar plot of S21 is shown in Figure 11

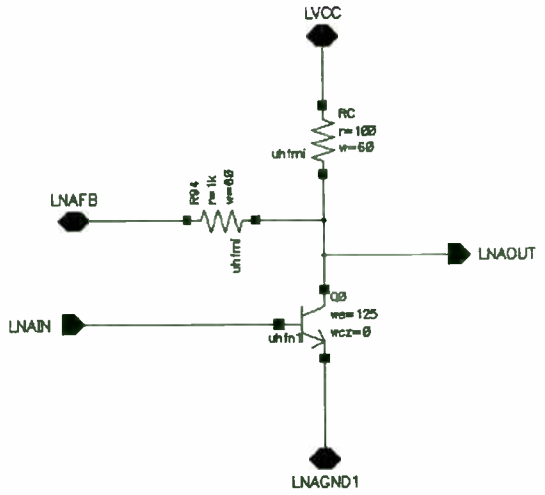


FIGURE 9. Schematic of a low noise amplifier circuit.

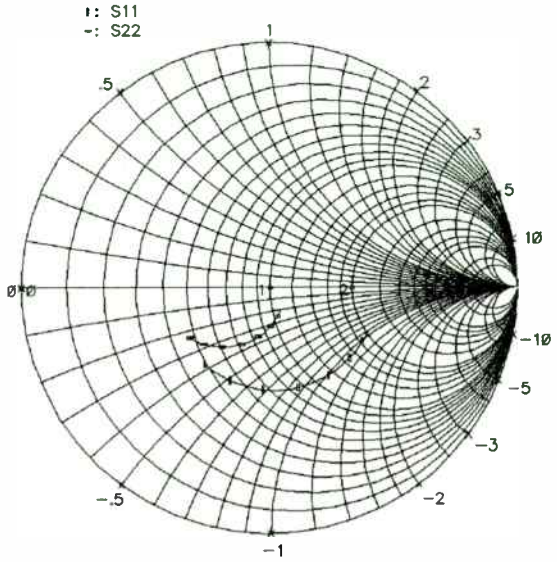


FIGURE 10. Plots of S11 and S22 on a Smith chart. The frequency varied from 100MHz to 1GHz.

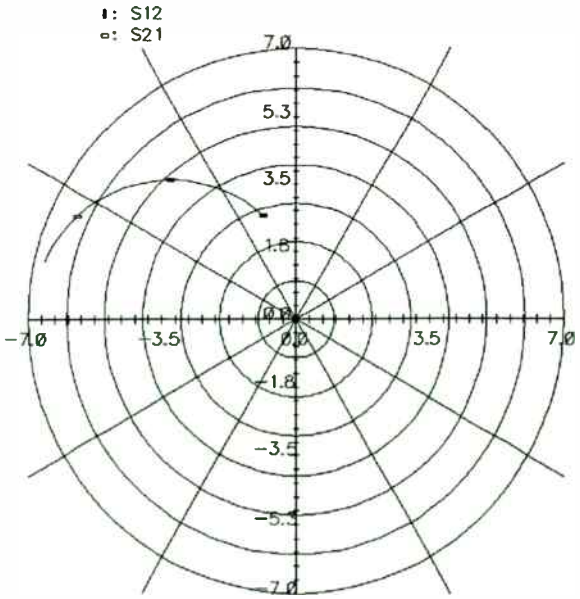


FIGURE 11. Polar plot of S21. The frequency varied from 100MHz to 1GHz.

The S parameters were then used to construct constant gain and input/output stability circles (shown in Figure 12). Note that portions of the stability circles are plotted in the upper left-hand corner of the figure. None of these analyses required the user to add any additional circuitry to the schematic and all plots were generated by FIRST. The user was not required to transfer or convert data in any way.

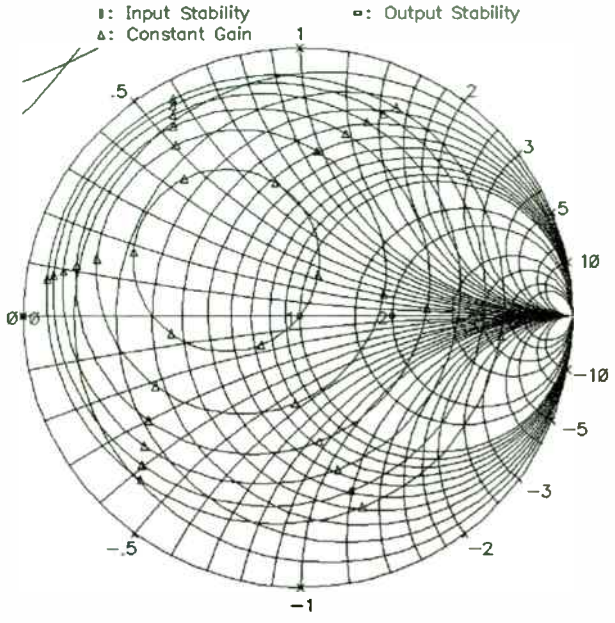


FIGURE 12. Constant gain and stability circles plotted on a Smith chart at a frequency of 1GHz. Gain circles are plotted at 1 dB intervals.

The noise analysis tool in FIRST was utilized to obtain the noise optimization parameters (minimum noise figure, optimum reflection coefficient, and noise resistance).

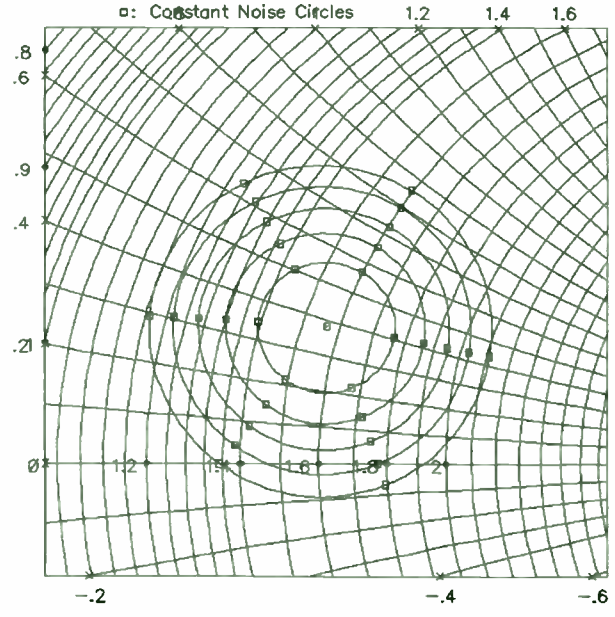


FIGURE 13. Zoomed-in view of noise circles plotted on a Smith Chart at 1GHz. Noise circles are plotted at increments of 0.5 dB.

This information was used to construct the noise circles in Figure 13. Again, all post-processing steps, test circuit insertions and graphics generation are performed transparently.

With the addition of sinusoidal input sources and an output load to the schematic the user has a choice of two forms of distortion analysis. For high accuracy at given input frequencies a large signal AC analysis can be performed (Figure 14). Note the high dynamic range of the frequency spectrum.

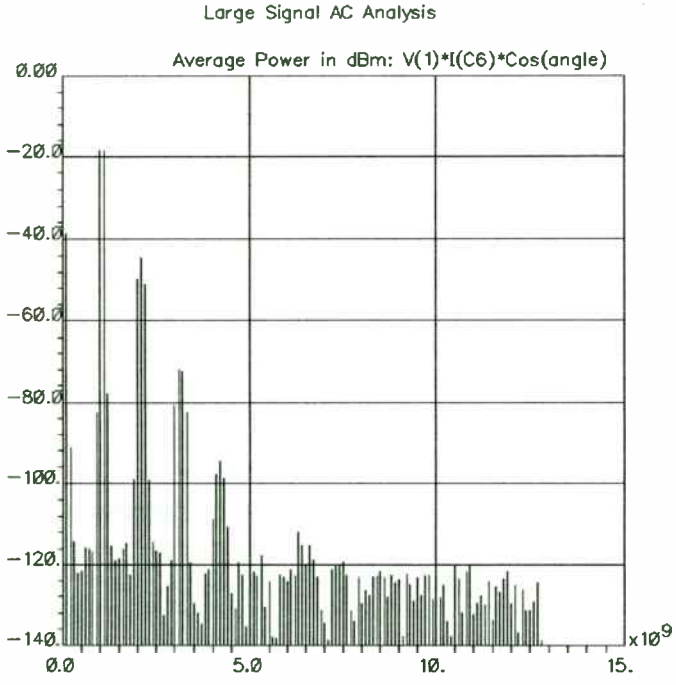


FIGURE 14. Magnitude of large-signal AC response of LNA circuit with input sources operating at 1GHz and 1.1 GHz.

Alternatively, a small signal distortion analysis can be performed when the desired information is the power of the distortion components over a range of fundamental frequencies. Figure 15 shows a plot of the output third-order intercept point as a function of the fundamental frequency. Note that both the large signal and small signal analysis show an output third order intercept point at around 10 dBm.

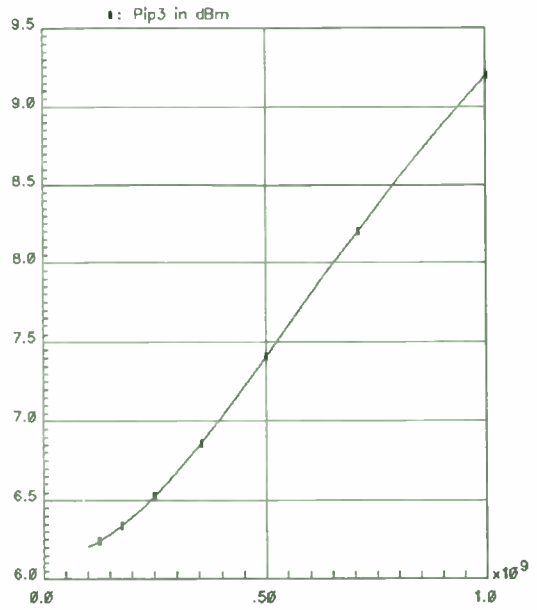


FIGURE 15. Plot of the third-order intercept point over the fundamental frequency with one input source operating at 0.9091 times the frequency of the other.

Once performance specifications were met, the layout of the LNA/Mixer was constructed using the FASTRACK layout tools described earlier (Figure 16). These tools were also used to verify the physical and electrical integrity of the layout. At this stage interconnect parasitic resistance and capacitance can be extracted and automatically back-annotated to the schematic netlist for post-layout simulations.

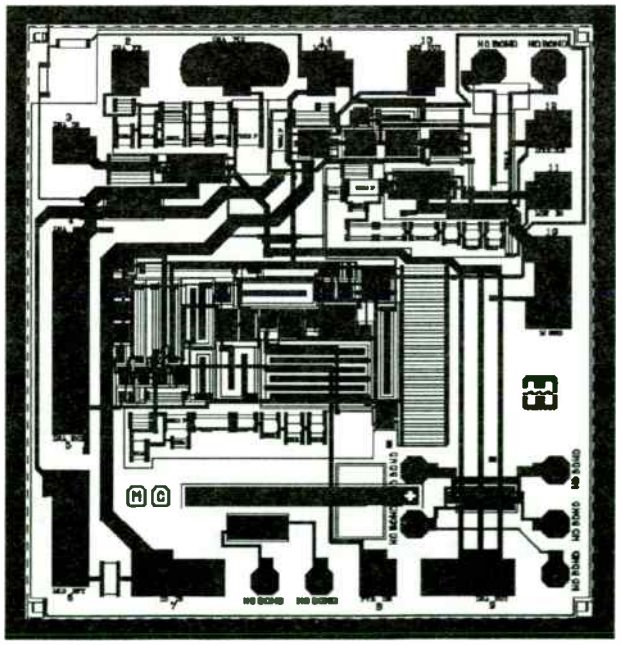


FIGURE 16. Layout of a combination LNA/Mixer circuit.

V. Conclusions

Embedding RF tools in an IC design system is an enabling factor for the transition from discrete based designs to using silicon ASICs. FIRST is a set of embedded RF design tools in Harris Fastrack's IC design system. It enables high frequency IC designers to easily traverse to the RF domain and it provides a traditional RF- design-system-like environment for RF designers. Regardless of IC or RF design, the user interface, simulation engine, and device models are the same. FIRST has been successfully used by many HSS' internal and external high frequency IC and RF designers.

VI. References

[1] P. R. Gray and R. G. Meyer, Analysis and Design of Analog Integrated Circuits, Second Edition, John Wiley & Sons, New York, 1984.

[2] G. Gonzalez, Microwave Transistor Analysis and Design, Prentice-Hall, Inc., New Jersey, 1984.

[3] R. S. Carson, High Frequency Amplifiers, Second Edition, John Wiley & Sons, New York, 1982.

[4] "Harris Fastrack Design System", Harris Semiconductor, Release 3.6, Sep. 1993.

[5] "Cadence SPICE Reference Manual", Cadence Design Systems, Version 4.2.2, June cdsSpice

[6] L. W. Nagel, "SPICE2, A Computer Program to Simulate Semiconductor Circuits", Technical Report ERL-M520, UC Berkeley, May 1975.

[7] R. A. Anderson, "S-Parameter Techniques for Faster, More Accurate Network Design", Hewlett-Packard Journal, Vol. 18, No. 6, Feb. 1967.

[8] T. B. Mills, "S-Parameters in Spice", RF Design, June, 1989.

[9] J. Ortiz and C. Denig, "Noise Figure Analysis Using Spice", Microwave Journal, April 1992.

[10] P. Antognetti and G. Massobrio, Semiconductor Device Modeling with Spice, McGraw-Hill Book Company, New York, 1988.

[11] S. Chilsholm and L. Nagel, "Efficient Computer Simulation of Distortion in Electronic Circuits", IEEE Trans. Circuit Theory, 1973

"Electromagnetic Simulation for High-Frequency Planar Circuits"

Daren McClearnon
Hewlett-Packard
1400 Fountaingrove Parkway, M/S 2US-P
Santa Rosa, CA 95403

Abstract

Electromagnetic (EM) simulation is emerging as an important way for high-frequency circuit designers to enhance the performance of their circuit-oriented simulation tools. This paper begins by identifying the role of Method of Moments (MoM) EM simulation among high-frequency simulation tools. It explores the advantages of using a multilayer planar EM simulator (based on a recent proprietary implementation of the MoM analysis technique) to supplement traditional circuit simulation tools. The planar Method of Moments approach is contrasted with a more general 3-dimensional approach using the Finite-Element Method and is shown to have significant advantages for most circuit designers.

Introduction

Electromagnetic simulation is outgrowing its reputation as an expensive, time-consuming, and exclusively academic endeavor. Electromagnetic simulation tools are now part of the standard product lines of all high-frequency CAE vendors. They have become integrated within CAE environments for schematics and artwork, enjoying better user-interface design, documentation, and support than ever before. Electromagnetic simulation is "here" today.

This is an informal discussion of today's state-of-the-art in high-frequency electromagnetic simulation. The discussion is oriented toward circuit designers and engineering managers with an eye toward the practical benefits of electromagnetics, rather than a detailed look at the underlying principles. Recently, Hewlett-Packard introduced a MoM-based simulator; it is used to illustrate current trends in the industry and electromagnetics. The simulator is called HP Momentum. It is considered a "2.5-dimensional" simulator since it assumes a planar circuit, but allows many planes to be entered and connected with vias. Specific product information is available directly from Hewlett-Packard and is not discussed here.

The Need for Greater Accuracy

Although commercial electromagnetic simulators for the high-frequency market have come a long way in the last five years, this discussion wouldn't be necessary if normal circuit simulators completely "did the job."

Circuit simulators operate within a framework of assumptions and simplifications that reflects the way electrical engineers think about circuits: in terms of schematics that show connections between components. Schematics are abstract, symbolic; they are only related to a physical implementation to the extent that a designer includes the extra details (*i.e.* - *"physical modeling"*). This level of detail includes parasitics, self-resonances, microstrip transmission lines and discontinuities.

One key assumption is that each of these second-order effects is independent of everything else in the circuit. Depending on the frequency range and the physical distance between phenomena, this is more than just a valid assumption; it makes circuit simulation *feasible*. A completely interconnected circuit would be too hopelessly complex to understand or simulate in any reasonable time.

However, as frequencies and densities increase, certain weaknesses in the assumptions begin to appear. A few of the possible interactions become significant, such as coupling between adjacent traces, or non-ideality of the ground plane. This then is the justification for using some form of electromagnetic analysis: to enhance the accuracy of an ordinary circuit simulation over an extended range of topologies, materials, and frequencies.

In some situations, the electrical representation of a structure itself changes with increasing frequency, from an inductance to a high-impedance transmission line to a set of n-coupled striplines in an IC package. In other situations, circuit features begin to interact through radiation, higher-order modes, or surface waves.

E&M simulators can quantify whether a certain phenomenon is significant and allow the designer to enter the result directly into a circuit, all from within the circuit-oriented CAE environment. The established path for this information is a linear S-parameter data file vs. frequency. Like the electromagnetic solvers themselves, S-parameters make no assumptions about whether crosstalk or other phenomena are best described in terms of lumped elements, distributed elements, or fullwave formulations. The S-parameters are simply inserted into the circuit at the location of the interaction without interpretation. This means that design engineers can reliably obtain accurate answers with a minimum of sophistication in the field of electromagnetics.

Physical Phenomena Not Ordinarily Modeled in Circuit Simulators

A variety of components and phenomena that circuit simulators do not address are listed in this section. Due to time constraints, it is not possible to elaborate in detail on all the items listed below. A small sampling of real circuits is shown to illustrate the principles.

Physical phenomena and structures not ordinarily modeled by circuit simulators

Phenomenon or structure	Example(s)	Illustration
Proximity, coupling and crosstalk effects	Coupled-line filters, TAB leads in a digital pkg.	Figure 5
Higher-order and evanescent mode effects	Coupled microstrip discontinuiites	Figures 4,6
Nonstandard structures with no commercially implemented model	Microstrip Y-junctions, planar baluns	---
Discontinuous or optimized ground planes	Slotted ground planes, cut-outs under spiral inductors, digital package power/ground planes	Figure 3
Transitions between propagating media	Microstrip to buried layers or coplanar waveguide	---
Radiation and surface wave phenomena	Microstrip patch antennas	Figures 1, 2
Wide variety of dielectric materials/thicknesses	Models used beyond accuracy guidelines	---
Cover heights and passivation layers	Package lids, buried microstrip	---

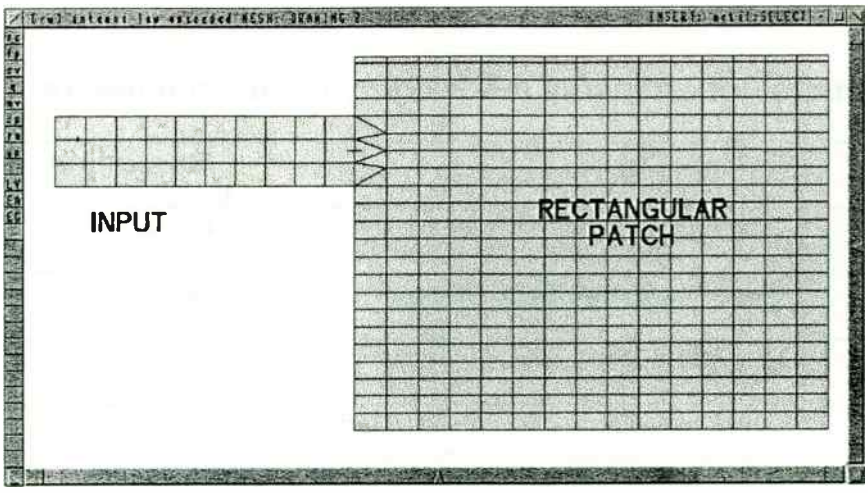
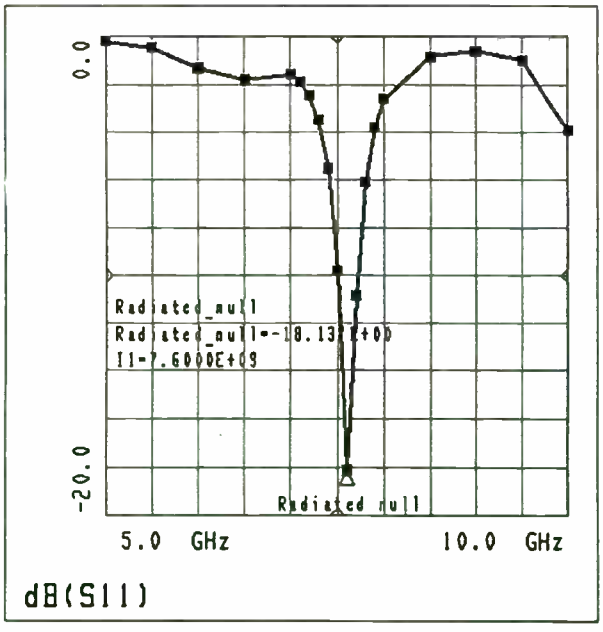
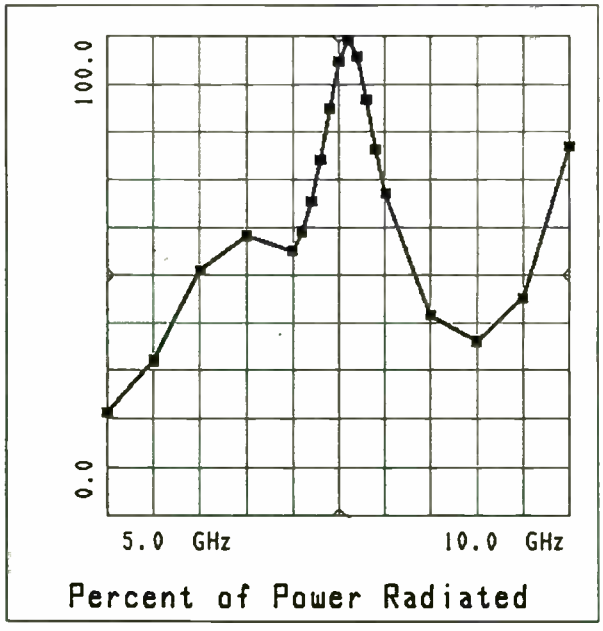


Figure 1. *Rectangular microstrip patch antenna.* Note that the magnitude of S_{11} varies significantly, even though the structure has only one port and themetallization is relatively lossless. This indicates that much of the energy of the system is lost to radiation. This effect is not modeled by circuit simulators.





The measurements for this example were obtained from a paper in the IEEE Transactions on Antennas and Propagation.

Shih-Chang Hu, N. Alexopoulos, D. Fordham, "Feeding Structure Contributions to Radiation by Patch Antennas with Rectangular Boundaries", IEEE Trans. on APS, vol. 40, no. 10, pp. 1245-49, October 1992.

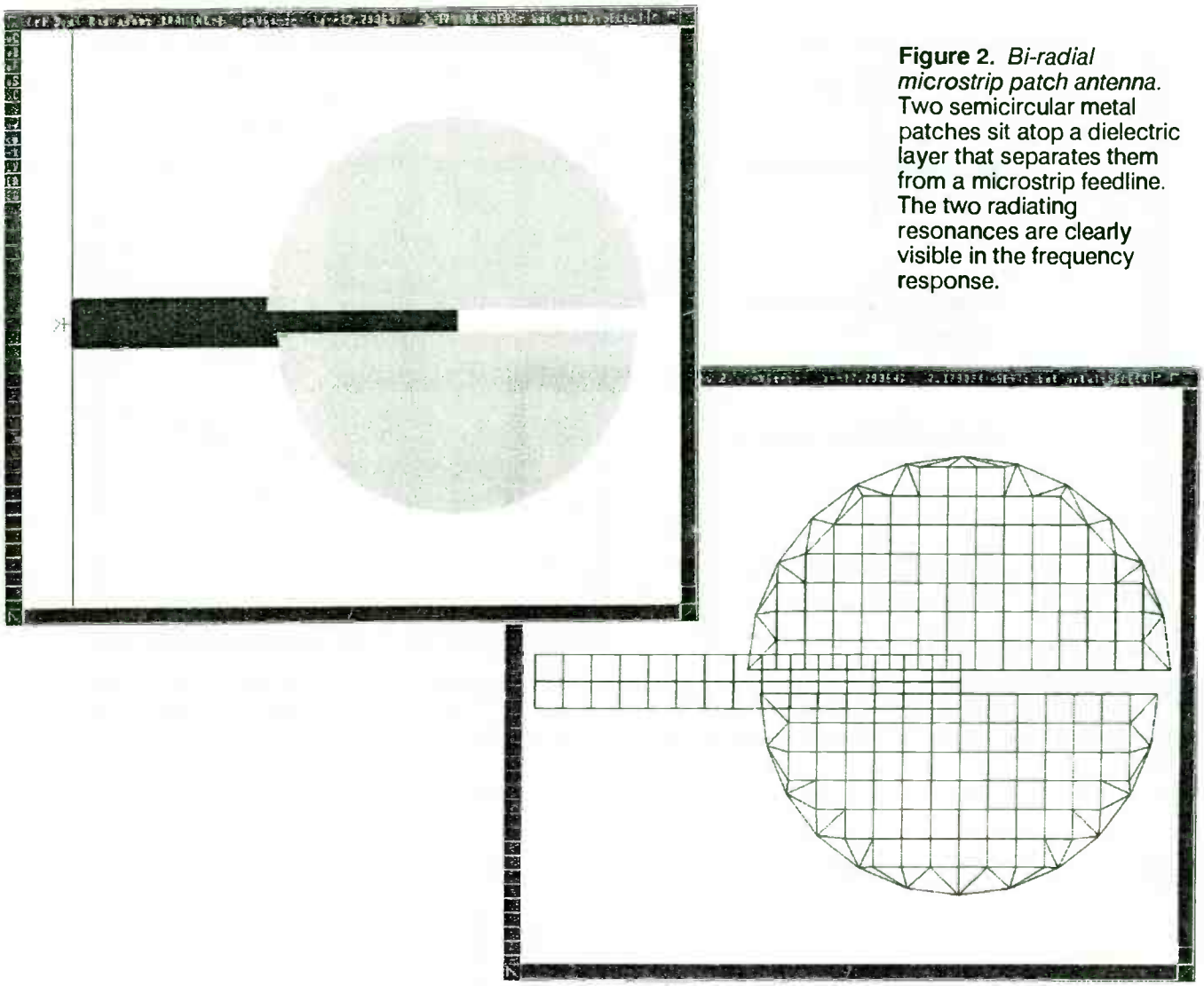
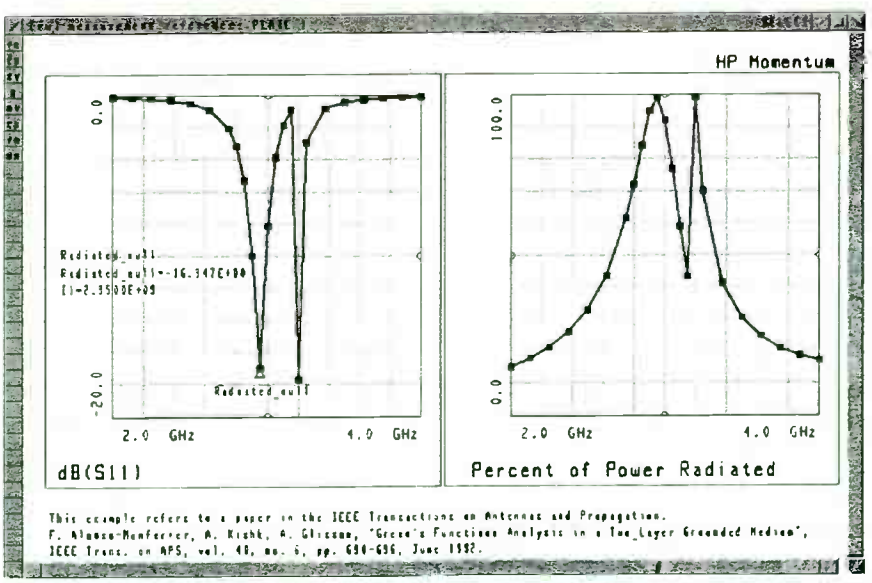


Figure 2. *Bi-radial microstrip patch antenna.* Two semicircular metal patches sit atop a dielectric layer that separates them from a microstrip feedline. The two radiating resonances are clearly visible in the frequency response.



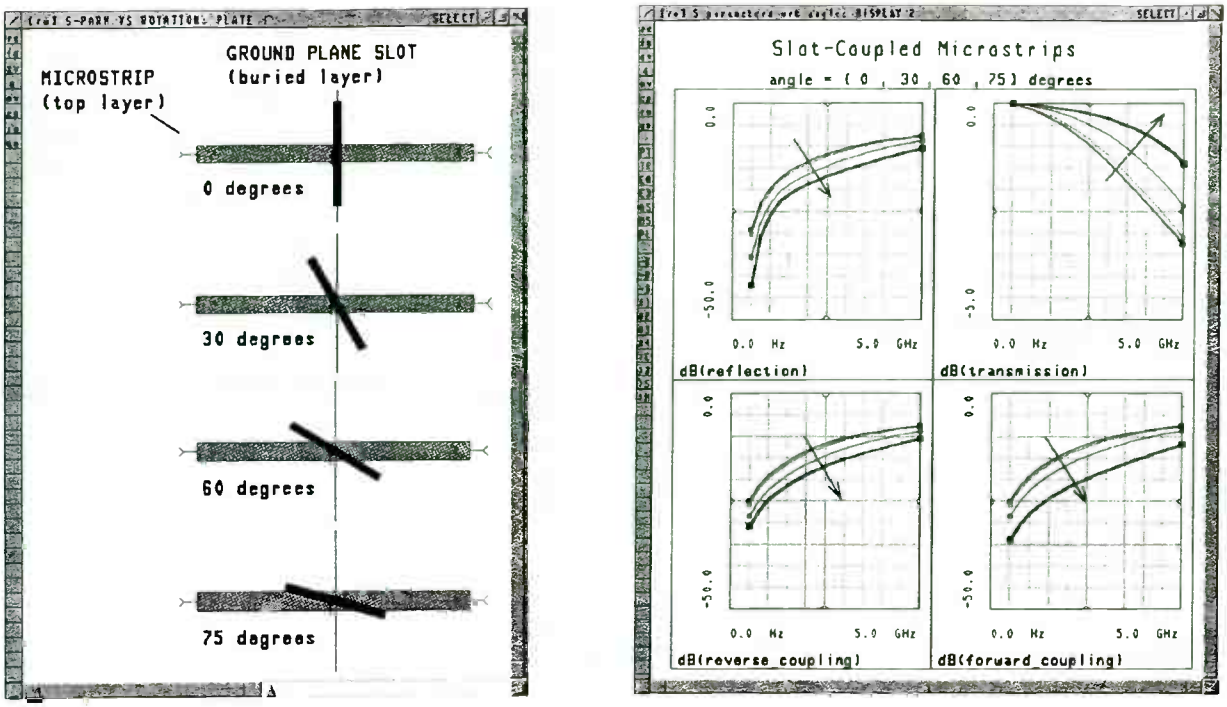


Figure 3. *Slot-coupled microstrip lines.* A two-layer dielectric has a ground plane sandwiched in the middle. Microstrip lines are established above and below the ground plane. They are coupled through a slot in the ground plane that has been rotated at various angles from perpendicular.

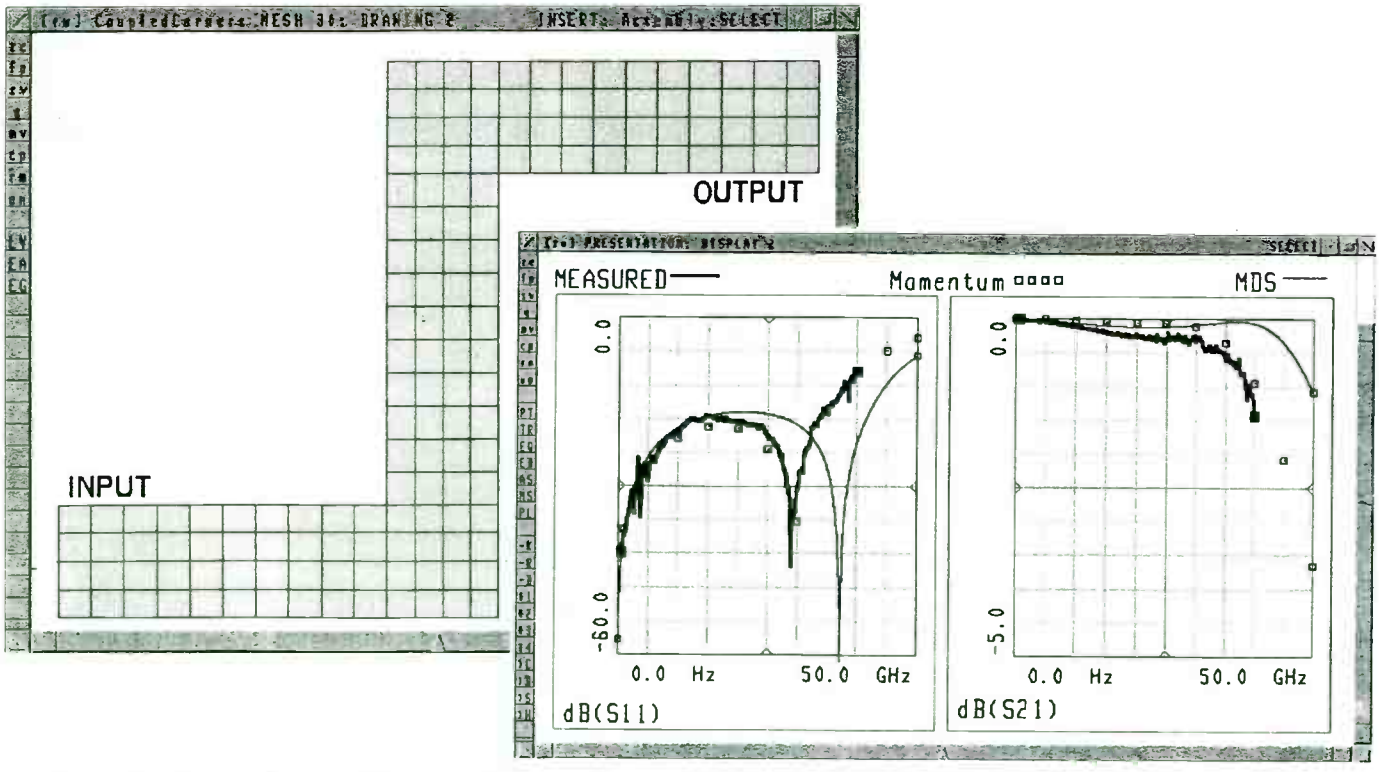


Figure 4. *Coupled microstrip corner discontinuities.* Even though these two simple microstrip corners are separated by a distance of several substrate-heights, they are coupled through higher-order modes excited in their vicinity. In practice, this alternate propagation path shifts a resonance down in frequency, a subtelty predicted by a MoM simulator and verfified by measurements, but not modeled by a circuit simulator.

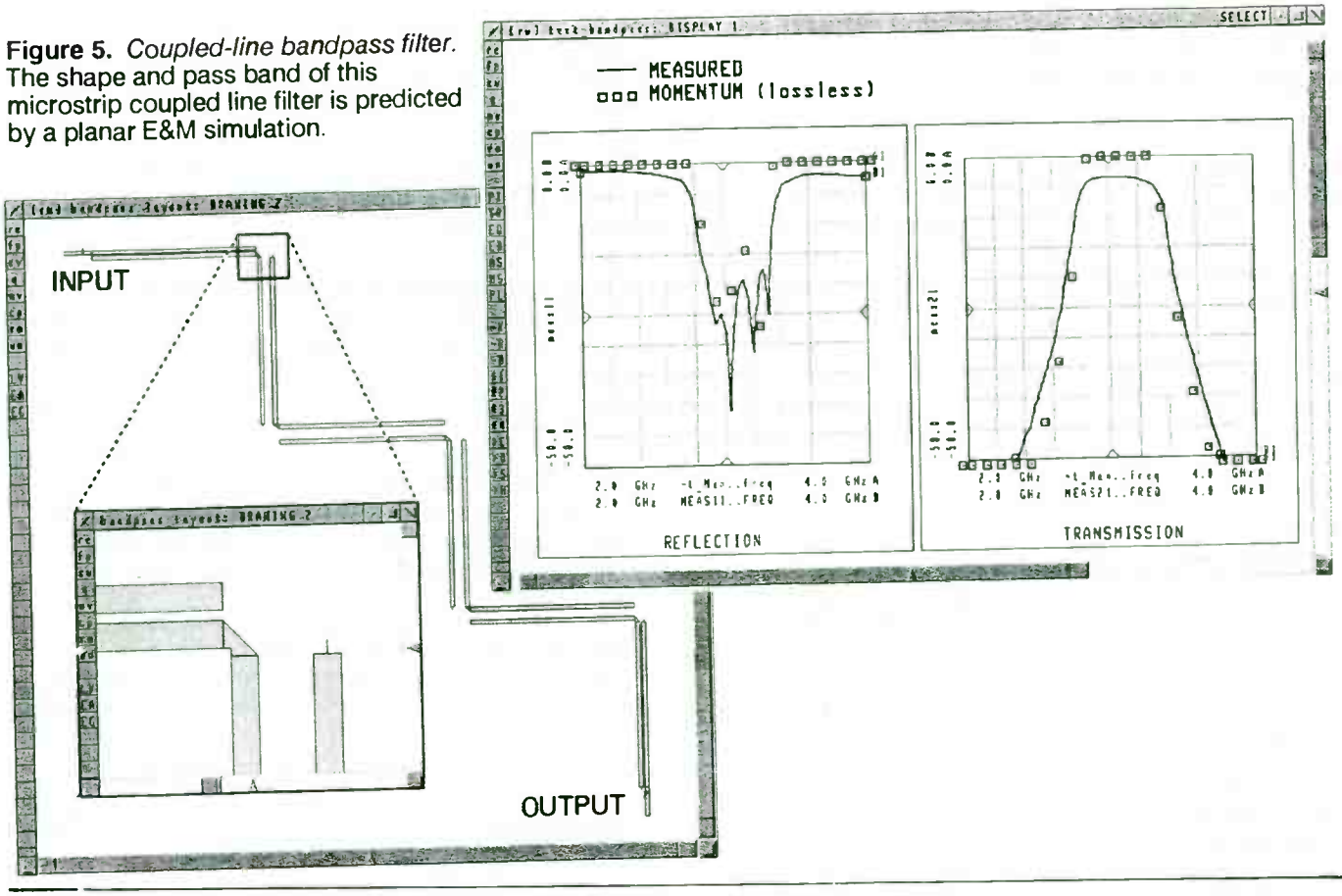


Figure 5. *Coupled-line bandpass filter.* The shape and pass band of this microstrip coupled line filter is predicted by a planar E&M simulation.

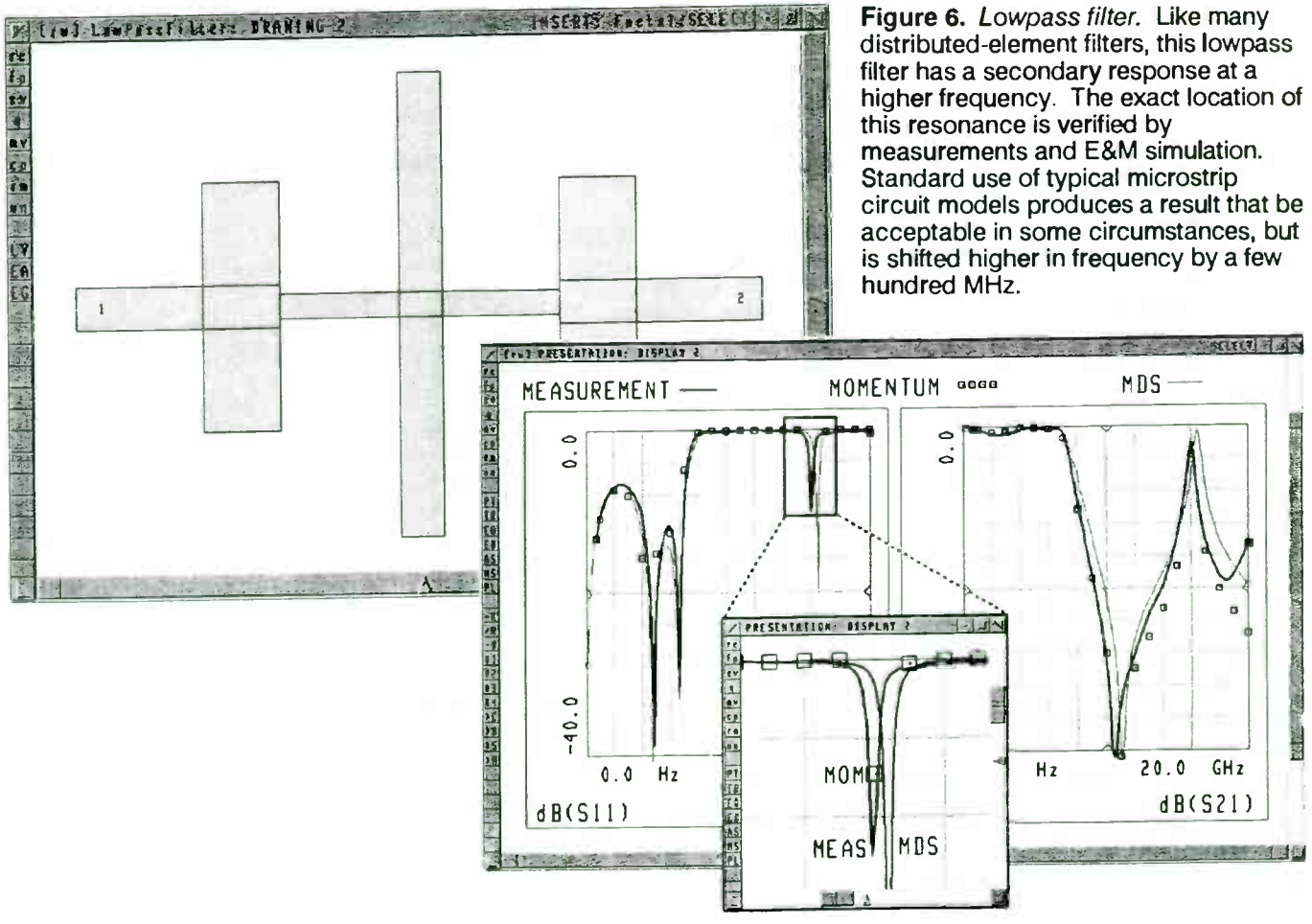


Figure 6. *Lowpass filter.* Like many distributed-element filters, this lowpass filter has a secondary response at a higher frequency. The exact location of this resonance is verified by measurements and E&M simulation. Standard use of typical microstrip circuit models produces a result that be acceptable in some circumstances, but is shifted higher in frequency by a few hundred MHz.

Comparison of different EM simulation methods

There are two dominant electromagnetic solution techniques in commercial use today, Method of Moments and the Finite Element Method. The basic differences between the two approaches are illustrated and summarized here using two commercially-available simulators from Hewlett-Packard. They are the 2.5-dimensional (multi-layer planar) MoM simulator (HP Momentum) and full 3-dimensional FEM simulator (HP HFSS).

3D Finite Element Method

The 3-dimensional approach is the more general of the two alternatives discussed. The working equation used by HP HFSS is an energy minimization problem shown below.

$$F(\bar{E}) = \int_{\Omega_d} \{(\nabla \times \bar{E})^2 + k^2 \bar{E}^2\} d\Omega + \sum_{i=1}^{2} \int_{Pi} (\hat{n} \times \nabla \times \bar{E}) \cdot d\bar{S}$$

Equation 1. *The HFSS energy equation*

First, the volume is subdivided into a 3-dimensional mesh. The port solutions are then determined and used as "excitations" on the structure. The energy within the volume is minimized and the electric field that satisfies Maxwell's Equations is determined. The S-parameters are then calculated from the resulting field distributions.

In the equation above, the volume and the port definitions are known, leaving the electric field (E vector) as the only unknown. The problem statement makes few assumptions about the nature of the circuit and allows for complete generality.

2.5D Method of Moments

The 2.5-dimensional approach assumes that the circuit is planar, or a series of planes connected by idealized vertical currents (vias). It also assumes that the metallization has zero thickness. For circuit designers, these operating assumptions are quite reasonable. In return for these assumptions, the problem can be stated as shown in Equation 2.

$$\int_S dS' [G_m(r,r') J(r') - \nabla (G_e(r,r') \nabla' \cdot J(r'))] = -E^i(r)$$

Equation 2. *the Momentum equation.*

Green's Function

The above equation stands in contrast to Equation 1. `Ge()` and `Gm()` are known as the electric and magnetic Green's Functions. `J()` is the current distribution throughout the system and represents the only unknown.

Since the substrate dielectrics and thicknesses are known in advance, the Green's Functions can be calculated once and stored in a database. Although the pre-calculation of the Green's Functions may be a significant computational step, it is only required once. All subsequent simulations of any patterns of metallization on that substrate are greatly accelerated. In practice, the assumption of a planar circuit and pre-calculation of the Green's Function can reduce the overall computation time to seconds or minutes instead of minutes or hours (per frequency point).

Other Advantages of HP Momentum

HP Momentum has two additional advantages over the general 3-D FEM simulator. First, HP Momentum allows the user to choose between meshing the metal or absence of metal on a particular layer. This is useful for transmission media that use large ground planes, such as coplanar waveguide or slotline. It is also very efficient for estimating the effect of discontinuous ground planes. The 3-D FEM tool must create mesh elements in the entire volume, thus adding considerably to the size of the problem, which further impacts resource usage and solution times.

Finally, HP Momentum is an open-boundary simulator that does not require that the circuit to be placed inside a conducting metal box. This allows the simulator to account for radiation loss and surface waves correctly.

Disadvantages of Method of Moments

Like all good compromises, the efficiency of the Method of Moments technique comes at a price. First, there are large classes of electromagnetic problems that cannot be solved using a planar tool, the most common involving packaging and transitions between transmission media. Some multi-layer planar simulators have been represented as being quasi-3-D with the generous use of vias, but all commercial MoM simulators have similar limitations in this regard.

Growth of resource usage and solution times

Second, the Method of Moments formulation leaves behind a smaller, but full matrix to solve. In fact, this is a benefit for some circuits since the MoM technique tolerates a coarser mesh for comparable accuracy. However, the full matrix causes the resource usage to grow as the *square* of the number of unknowns, while the solution times grows as the *cube* as the number of unknowns. This is inescapably true of all of today's commercial MoM simulators.

In contrast, FEM matrixes remain sparse longer and grow at slightly lower geometric rate (closer to *n-squared*). At some point, MoM can lose its efficiency advantage and take longer than a full 3-D FEM simulation. This point seems to be on the order of several thousand unknowns but is difficult to verify, since the crossover point occurs at the RAM and disk horizon of today's desktop computing capacity.

This crossover point could be reached sooner with inefficient gridding schemes. An example of an overly-conservative MoM mesh is the use of hundreds or thousands of tiny rectangles to simultaneously represent curved or angled patterns on a large pattern. A few triangular sections might be more than adequate for the situation.

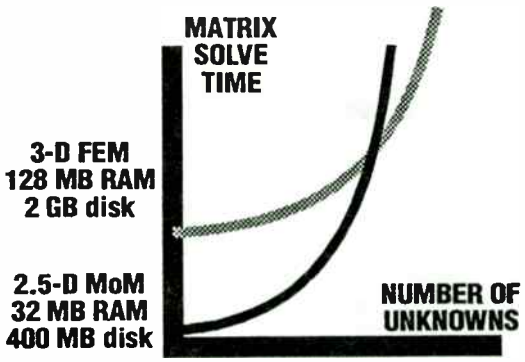


Figure 7. *Matrix solve time vs. number of unknowns*

Notice that in the nominal case, the 3-D FEM tool requires a dedicated "number-crunching" workstation for reasonable design capacity, while the MoM tool runs on the class of desktop workstation already used for typical CAE applications.

Conclusion

Many companies continue to invest in their own specialized electromagnetic tools in order to retain a proprietary advantage. However, as the commercial CAE vendors make each generation of these electromagnetic tools available to wider audiences, the tools are also being integrated more tightly into the mainstream CAE environments. This makes the commercial tools more accessible, convenient, and typically makes them better-documented and supported relative to in-house tools.

Having taken this brief tour of electromagnetic simulation tools, the high-frequency circuit designer should feel confident that E&M tools have matured quickly to help meet the challenges of a new era. Method of Moments simulators are of particular value to circuit designers because they provide accurate results quickly from within the CAE environment.

APPENDIX: Mechanics and Interaction with a MoM Simulator

For the benefit of designers who have never interacted with an E&M simulation tool, the mechanics and internal computations are summarized here at a high level. Principles are illustrated with screens and examples from HP Momentum, a recently introduced MoM-based simulator.

As viewed from the user's perspective, the amount of activity belies the actual complexity of events inside the software. At the very simplest level, the procedure has these steps (illustrated in figure 8):

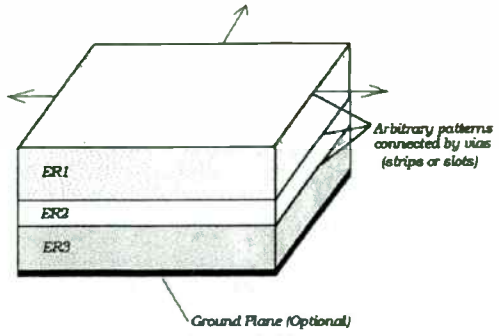
1. *Define a substrate and do a pre-calculation*
2. *Define the planar metallization and slot patterns, resistivities, and ports*
3. *Simulator subdivides pattern (a "mesh" is generated)*
4. *Feedline analysis is performed to determine port impedances and circuit parameters.*
5. *Method of Moments is applied to calculate distribution of currents, from which the S-parameters are calculated.*
6. *Results are fed back for plotting, visualization, and re-use in subsequent circuit simulations.*

The substrate definition and pre-calculation (Step 1) is done once and stored for re-use, while the software performs Steps 3 through 6 automatically. This leaves the drawing of the planar patterns (Step 2) as the only significant task for the user. In the case of HP Momentum, which is fed directly from the HP RF/Microwave Design System, this information comes directly from the schematic or layout that has already been entered for the circuit. Therefore, as much of the task as possible has been automated.

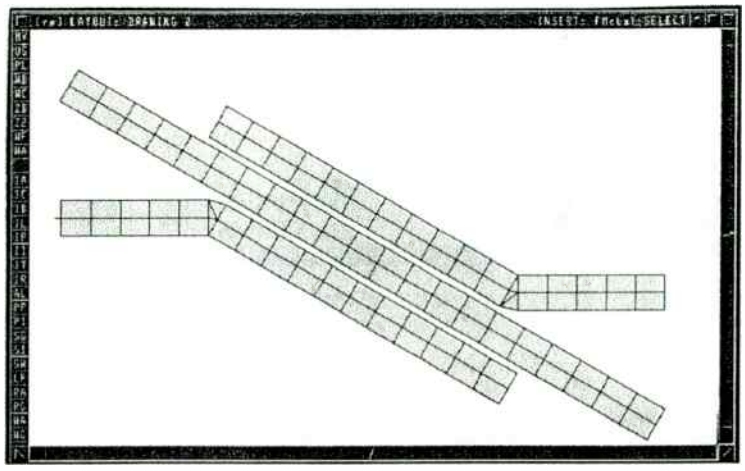
Acknowledgements:

The author wishes to thank Alex Anger, Brad Brim, Charles Plott, and Dave Wilson of Hewlett-Packard for their assistance in preparing this survey.

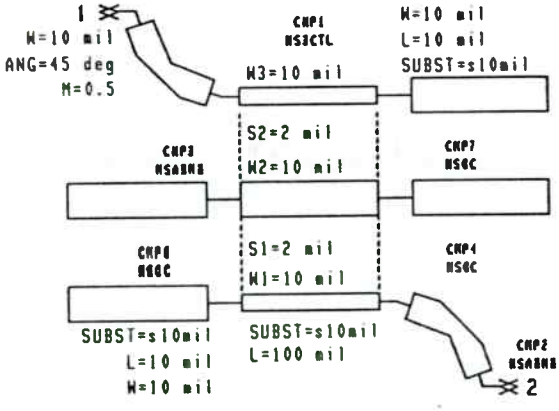
Figure 8. *Steps in Using a Method of Moments Simulator*

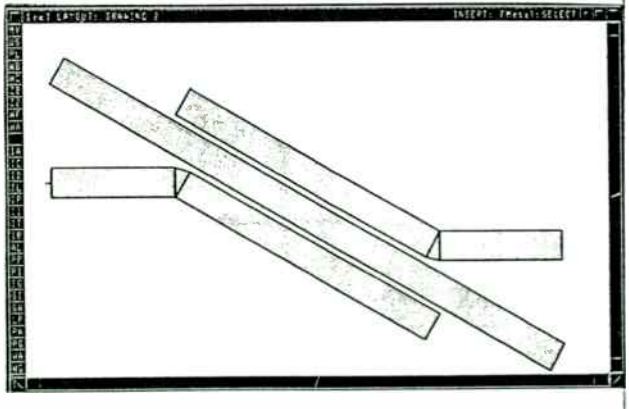


Step 1. Define the substrate and perform any precalculations.

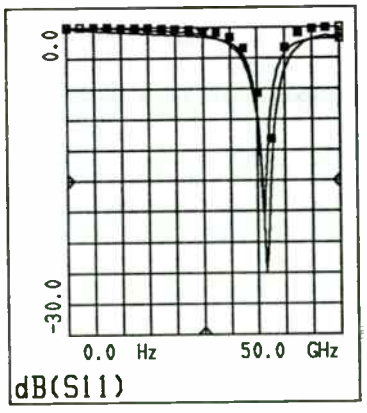


Step 3. Simulator subdivides structure into a "mesh". *Step 4* (*not shown*) feedline analysis performed to determine port impedances and circuit parameters.

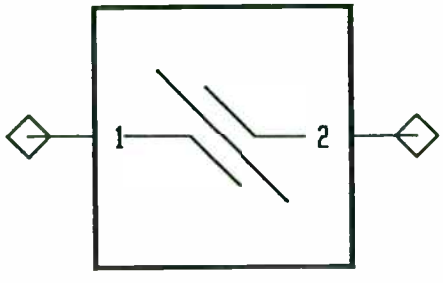




Step 2. Define the planar metallization and slot patterns and ports using schematic or graphical editors.



Step 5. Method of Moments is applied to calculate the distribution of currents, from which the S-parameters are calculated.



Step 6. Results are fed back for post-processing and re-use in circuit simulations.

A Monolithic 915 MHz Direct Sequence Spread Spectrum Transmitter

Stephen Press
Jeff Barry
Jim Marsh
Tektronix Microelectronics
P.O. Box 500
Beaverton, Oregon 97077

The allocation of the Spread Spectrum bands has spawned a wide range of applications which demand high performance designs with economical implementations. This article describes one such design which is a 915 MHz Spread Spectrum transmitter implemented with a minimum number of parts and a low manufacturing cost. An overview of the system design will be presented, however the bulk of the discussion will concern the transmitter ASIC which is a monolithic implementation including the 915 MHz oscillator, a divide–by–eight prescaler, a pseudorandom code input buffer, an up–converting mixer, an AGC amplifier, and an output driver.

SYSTEM OVERVIEW

This design is part of a fault detection system that broadcasts a signal to a remote location, identifying the fault condition. The transmitted signal has to be successfully received at distances up to two miles. The sensitivity of the receiver requires the transmitted signal output be +20 dBm. The total transmitter design fits on a circuit board which is less than 1.5 square inches. As shown in Figure 1, there are two ASICs in the design, the bipolar ASIC described above and a CMOS ASIC. The CMOS ASIC includes a pseudo–random code generator, a crystal controlled oscillator, a phase detector, and an integrating amp. In addition to the two ASICs, the transmitter board contains a prescaler, a regulator, and various passive components.

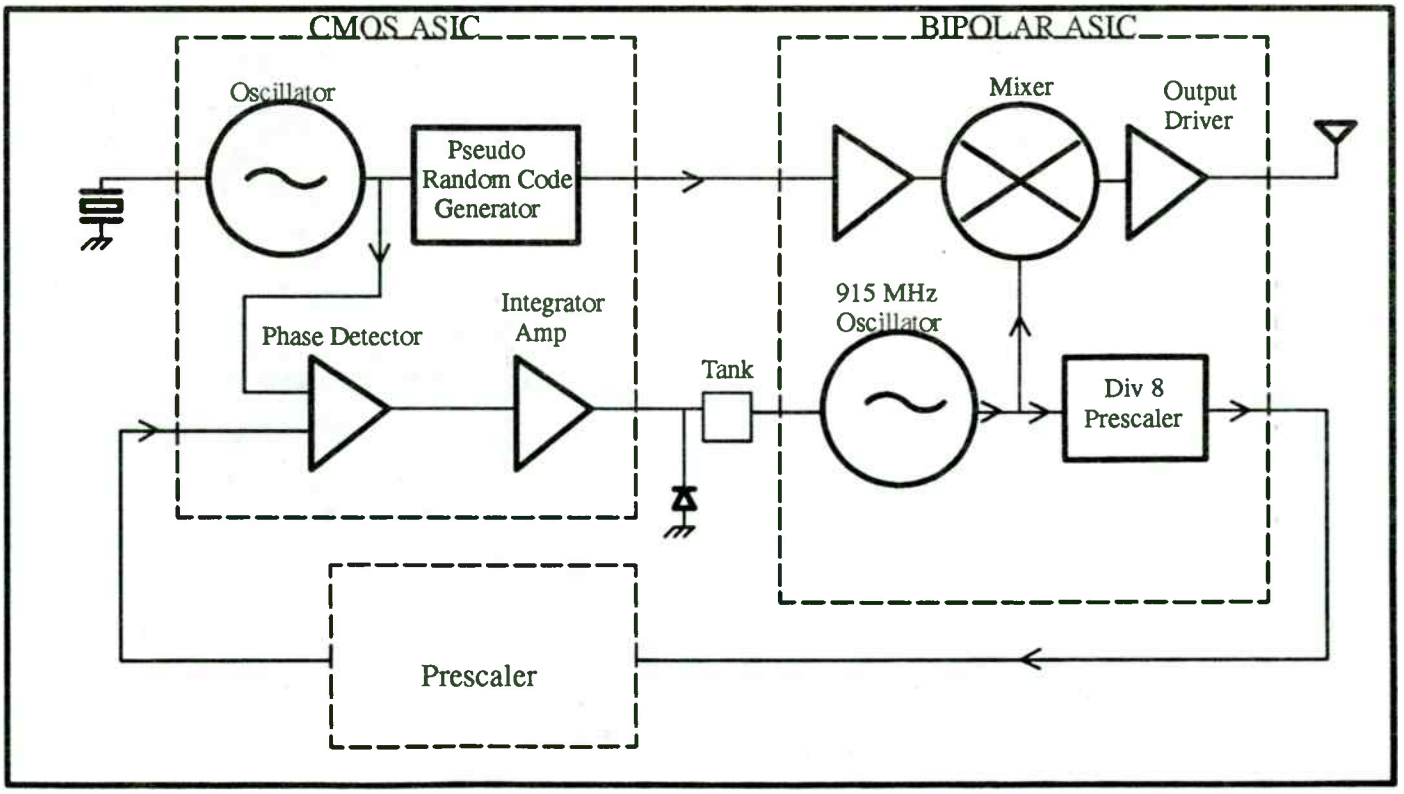


Figure 1. System Overview

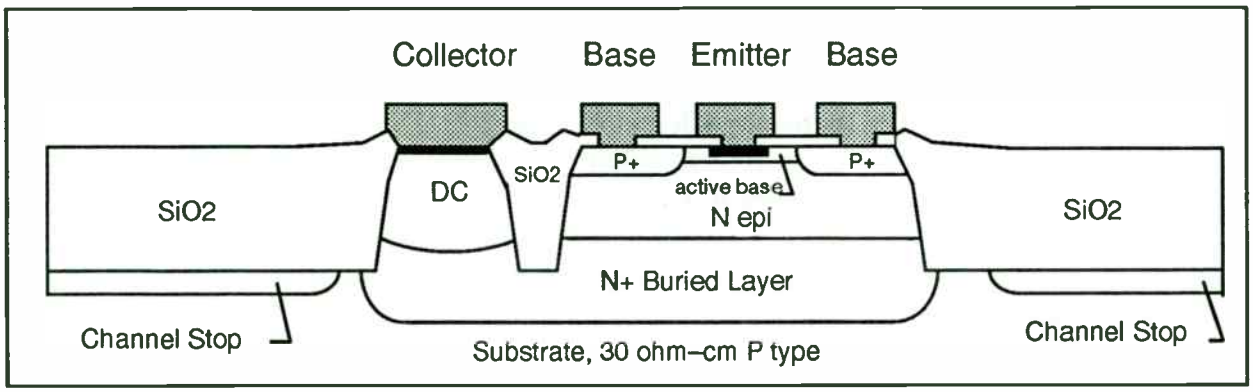


Figure 2. NPN transistor cross section

The crystal controlled oscillator on the CMOS ASIC provides the clock for the pseudo–random code (PN code) generator. The PN code is a 255 bit sequence which is encoded with information using Direct Sequence techniques. In this case there is one bit of information per complete sequence where logic 1 is the normal 255 bit sequence and 0 is the inversion of the sequence. The code is output from the CMOS ASIC to the Bipolar RF ASIC.

The 915 MHz oscillator on the bipolar ASIC is phase locked to the crystal controlled oscillator on the CMOS ASIC. The prescaler function is partitioned between the Bipolar ASIC and the external prescaler. The Prescaler output is compared to the crystal oscillator output by the phase/frequency detector (PFD) whose output remains linear over a range of $\pm 2\pi$ radians. The PFD output drives a high gain charge pump integrator. The filter network for the integrator is performed with discrete devices external to the CMOS ASIC. The voltage output from the integrator is fed back to a varactor diode which is part of the tank circuit in the 915 MHz oscillator.

In the bipolar ASIC the PN code is converted from CMOS logic levels to a differential signal with appropriate levels for the mixer. The 915 MHz oscillator output and the PN code are mixed, resulting in a several MHz flat band signal which is centered at 915 MHz. This signal is passed through a two stage output driver before the antenna. The antenna is a 37 ohm quarter wave monopole.

BIPOLAR TRANSMITTER ASIC

The Bipolar ASIC was fabricated on the Tektronix SHPi oxide isolated process [1] using two layers of metal interconnect and offering npn transistors with an F_T in excess of 9 GHz. A transistor cross section is shown in Figure 2. The design was executed on a QuickChip 6–40 which is a standard array of devices located in fixed positions on the die. The layout was accomplished by interconnecting devices using the two layers of metal. By choosing this standard array we were able to reduce fab time to 4 weeks, reduce the number of masks purchased to four, and significantly reduce the layout time. The QuickChip array provides vertical npn transistors, lateral pnp transistors, JFET transistors, Schottky diodes, MOS and junction capacitors, two types of implant resistors, and optional nichrome resistors. (Figure 3.)

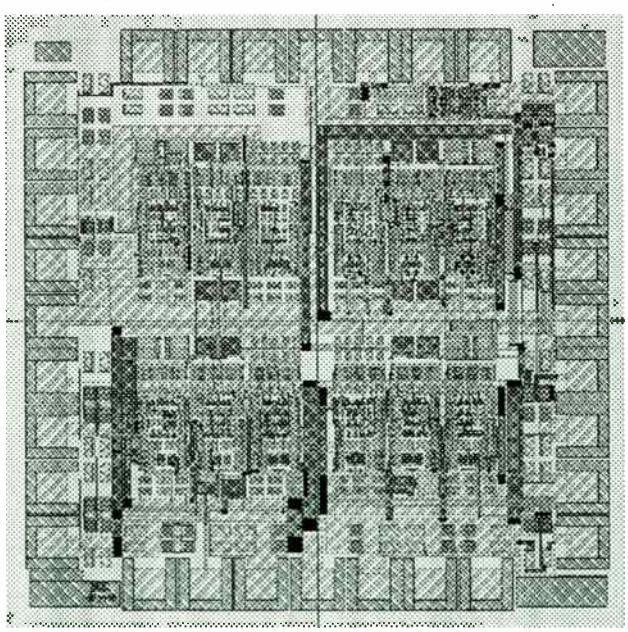


Figure 3. Bipolar Transmitter ASIC layout

An expanded block diagram of the Transmitter ASIC is shown in Figure 4.

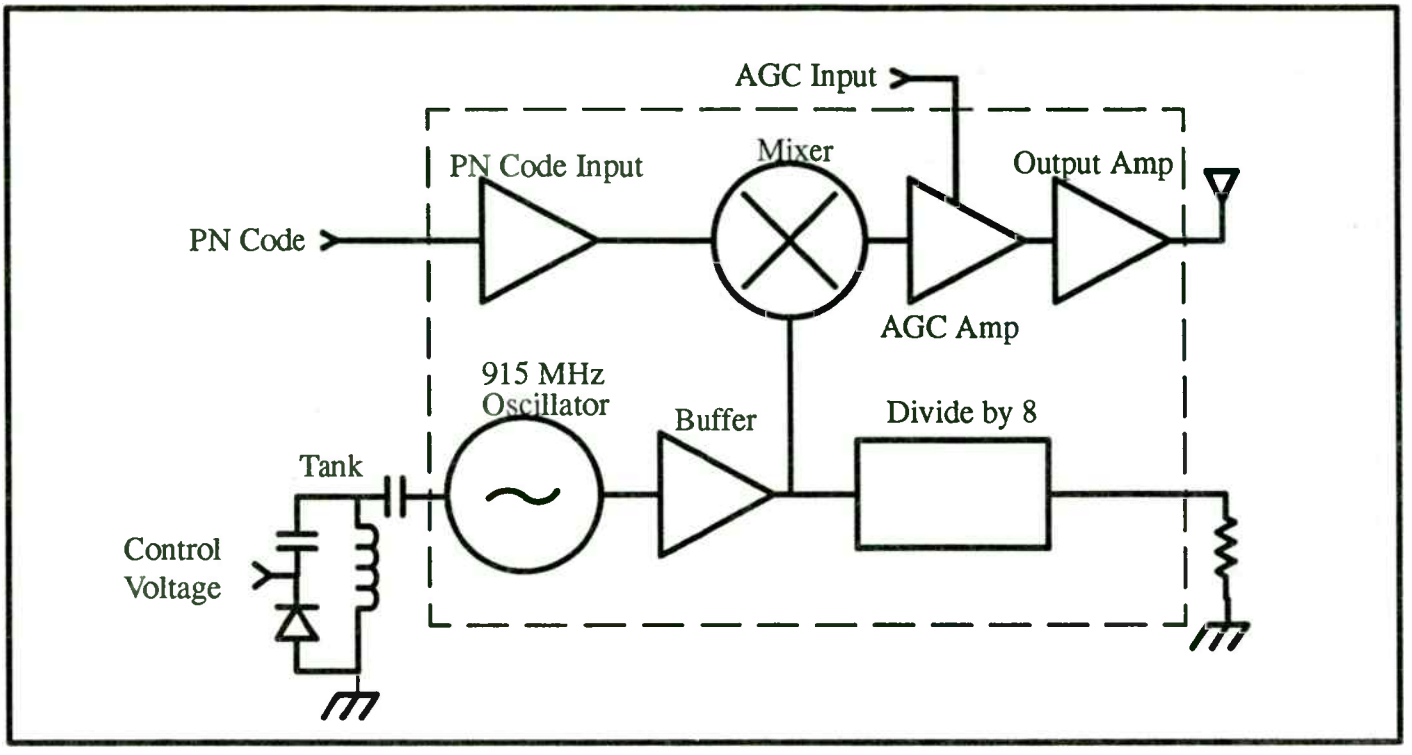


Figure 4. Transmitter ASIC

915 MHz OSCILLATOR

The negative impedance oscillator used in this circuit is single ended and resembles the Colpitts architecture. (See Figure 5.) The principal components include transistor Q7, capacitors C1 and C2, and the external tank circuit. In the tank circuit a varactor diode was used as a means of adjusting the frequency of oscillation. A parallel tank circuit was used to desensitize the oscillation frequency to bond wire inductance. In a negative impedance oscillator the real portion of impedance looking in from the tank is negative at the frequency of oscillation (and for some surrounding bandwidth). It is this negative impedance that allows the circuit to oscillate. The reflection coefficient looking into a negative impedance is more than one, which implies a continually growing oscillation at the resonant frequency of the tank. The factor limiting the amplitude of oscillation varies in different oscillator architectures. However a convenient way to view amplitude limits can be presented with a series resonant tank. In a series resonnant tank there will be a parasitic resistance in the L and the C tank elements and the trace connecting the elements and the oscillator on the IC. The relationship between this resistance and the inductance and capacitance of the tank can form the main contribution to determining the Q of the circuit where;

$$Q = \frac{1}{R} * \sqrt{\frac{L}{C}}$$

During a cycle of oscillation the sinusodal voltage variation changes the input impedance of the oscillator from the initial impedance at the DC bias point. This is due to large signal effects on the bias point of the transistor base. When the magnitude of the negative resistance of the oscillator equals the parasitic resistance then the oscillation amplitude will grow no further.[2][3]

The oscillation at the tank is transferred to the emitter of Q7 where it is converted to a current. This current is mirrored from Q8 to Q9. Q10 is a cascode stage used to increase the bandwidth. The mirrored current is converted to a voltage across R9. This signal is buffered by Q1 and fed into the single–ended–to–differential stage (Q2, Q3). R10 biases the undriven side of the differential stage (Q3). C3 provides an AC ground at this input. The signal is amplified across the differential pair, Q2 and Q3. One draw back to differential conversion using this technique is that the DC voltage dropped across R10 forms an offset on the Q3 side of the amplifier. To minimize the offset R10 must not be too large.

The initial amplitude of the oscillation at the tank is determined by large signal phenomena. Because it is difficult to predict the exact amplitude of the oscillation at the tank, the oscillator was designed to limit at the Q2–Q3 amplifier for all temperature and process conditions. Thus the limiting conditions of this amplifier, which are predictable, control the output amplitude of the oscillator stage. Limiting at this stage does introduce harmonics, however these are reduced by the bandwidth limit of later stages and by the matching network used at the output.

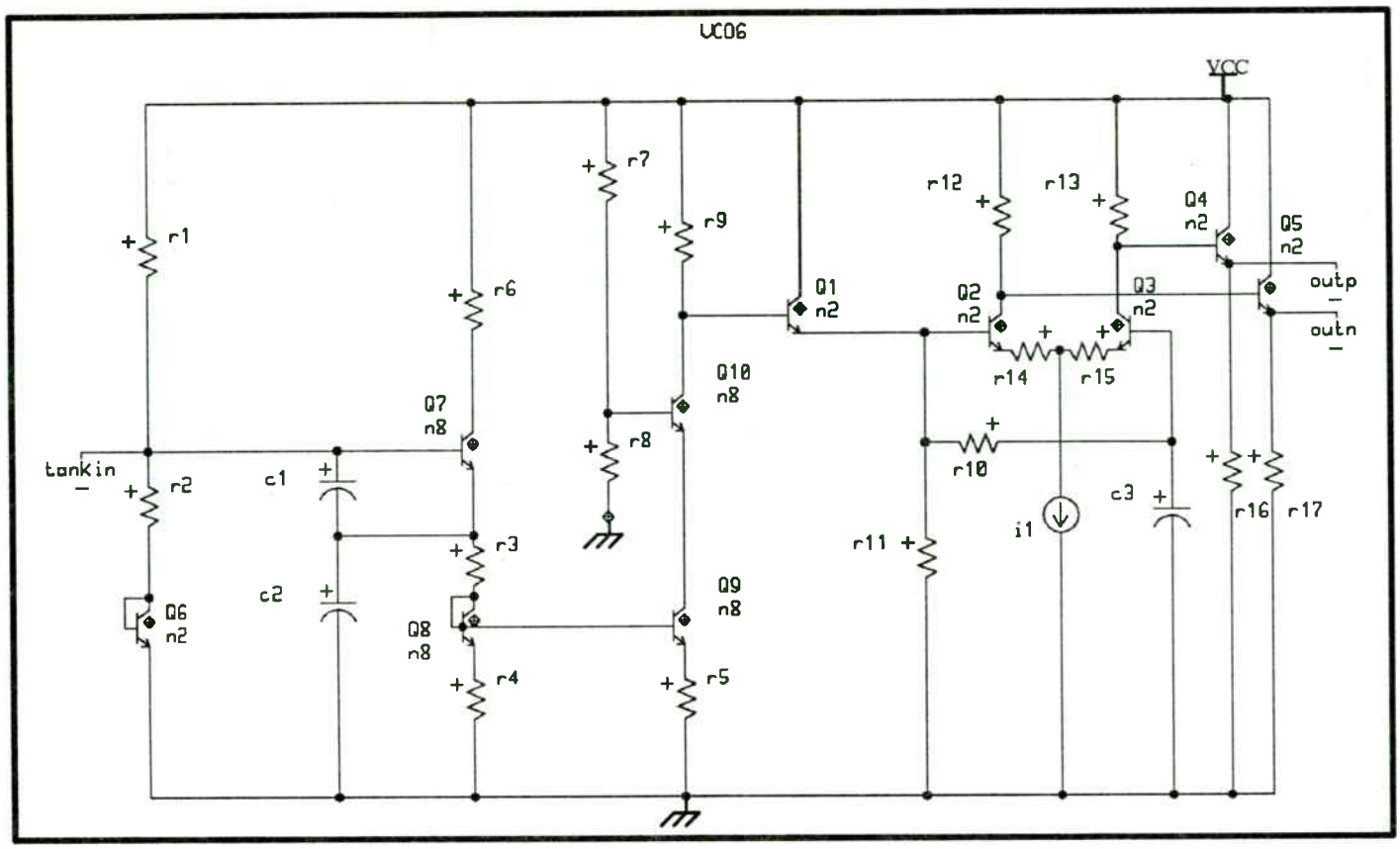


Figure 5. Simplified schematic of the VCO

THE PRESCALER

Although it would be desirable to put the complete prescaler on chip, this was not possible since this design was done on a standard array with limited numbers of devices. It is useful to have some of the prescaler on chip for two reasons. First it is an excellent buffer between the oscillator and the off chip prescaler and second, a 115 MHz signal coming off chip is handled much more easily than a 915 MHz signal.

The prescaler is a traditional ECL sequential divider using three flip–flops. Since the architecture is standard it will not be discussed further here.

PN CODE INPUT

The PN Code input is designed to accept a CMOS input signal. In this particular application the code rate is about 3 megabits per second (MBPS).

THE MIXER

The mixer as shown in Figure 6 is based on the Gilbert mixer design. Normally in a mixer the lower port (Q5 and Q6) is the RF port and the LO input is the upper port (Q2, Q3, Q7, and Q8). The amplifier formed with the lower differential pair is linearized with the addition of the emitter degeneration resistors R4 and R11. By so doing IP3 is increased and harmonics are decreased. Q5 and Q6 are large devices to lower noise on the RF port. However in a mixer which is part of a Spread Spectrum transmitter these rationales may not hold. As shown in Figure 6, the LO from the VCO is placed at the lower port and the PN code is placed at the upper port. This connection scheme is the opposite of that used in most applications. In a Spread Spectrum system the LO is phase modulated by the PN code, in particular the phase shift during a PN code transition is 180 degrees. The quicker this phase shift occurs the better defined is the information. The upper port accomplishes this task. The lower port accepts the high frequency 915 MHz LO. Because the lower port was used for the LO the bandwidth of the mixer remains relatively flat to 1 GHz. Two factors act to increase the bandwidth in the LO path through the mixer. First the transistors in the upper port, which are switching at a relatively slow rate (≈1.5 MHz), act as a cascode for the the transistors at the lower port. Second, by keeping the degeneration resistors (R4 and R11) at the lower port the gain is reduced to about 1.5 which also increases the bandwidth. The goal in this design was to keep the bandwidth flat past the frequency of transmission. As shown in Figure 7, simulation indicates this goal was achieved in the mixer.

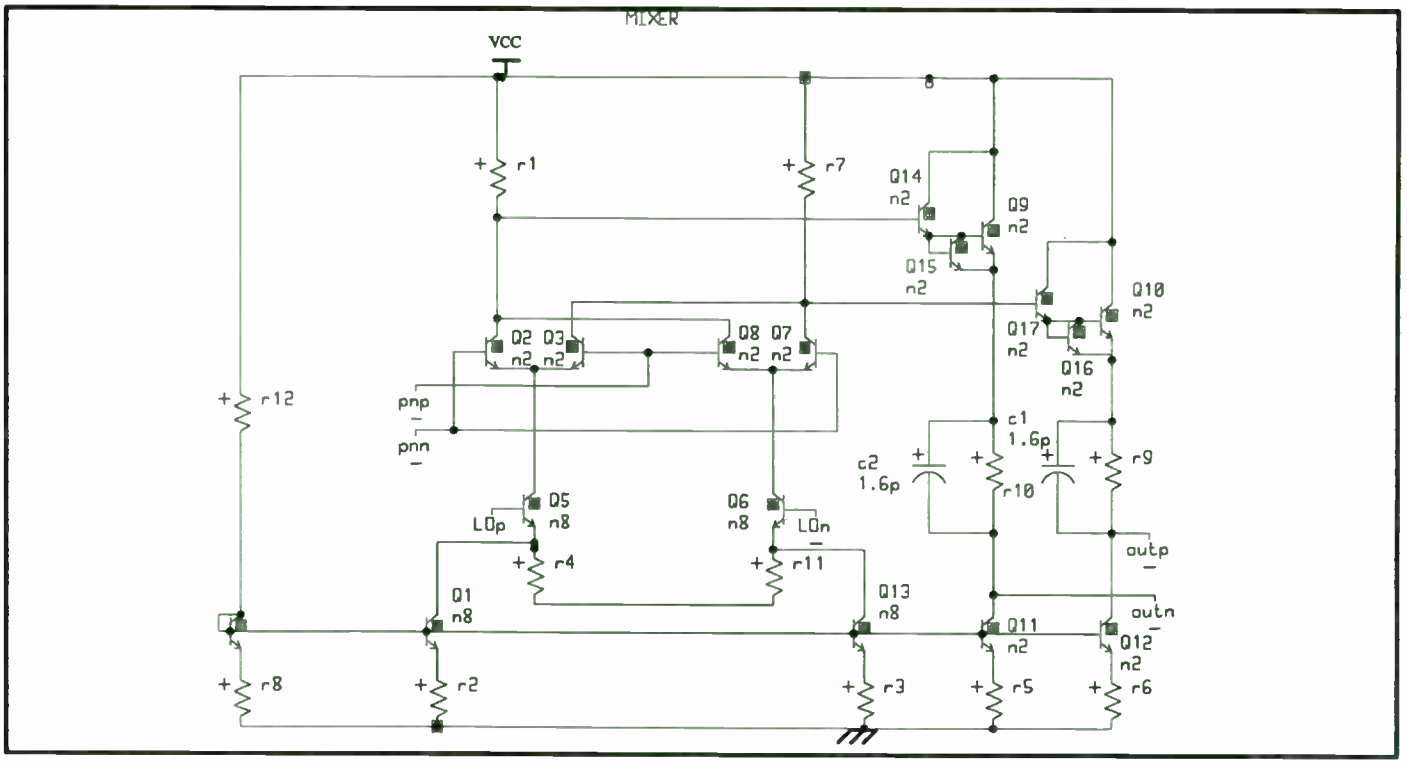


Figure 6. Simplified mixer schematic

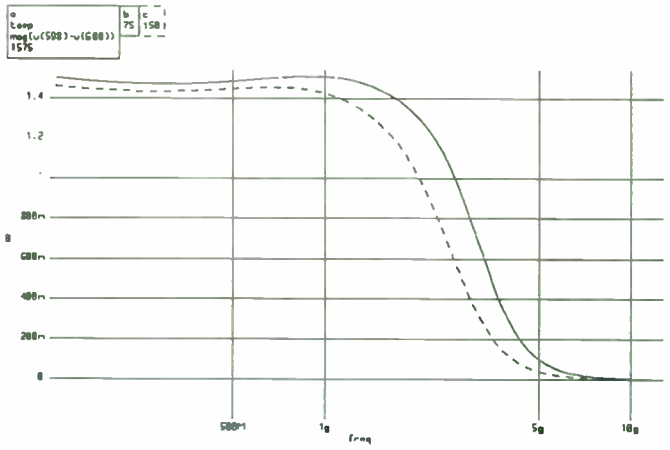


Figure 7. Mixer gain over bandwidth and temperature (simulated)

AGC AMPLIFIER

The peak to peak voltage coming out of the mixer is about 1 volt. The AGC amplifier increases this signal to about 1.8 volts. Since the circuit operates with a 5.2 volt supply and signal levels become relatively large, it is important to assure that for all signal conditions no transistors saturate. The mixer output is level shifted through two diode drops and an additional 1 volt through R9 and R10 before being passed on to the AGC amplifier. This allows the AGC enough headroom to swing a 1.8V signal. Amplification in the AGC stage is provided by the differential pair Q1 and Q2 (see Figure 8). Gain is adjusted by varying the bias current in Q1 and Q2. This is accomplished by circuitry consisting of Q7, Q3, Q4, Q12, Q13 and associated resistors. As the applied AGC voltage at the base of Q7 is reduced from the nominal 5.2 volts, the current supplied to Q1 and Q2 is reduced which decreases the gain. The AGC gain adjustment does not have to be linear for this system but it does requires 20dB of range which is achieved. (Figure 9). The purpose of the AGC is to cut back the transmitted power when the system battery power starts to fall. Since there is no requirement for a linear relationship between the AGC voltage and output power, the system described above is satisfactory.

The circuitry comprised of PLAT1, PLAT2, and Q16–Q20 keep the output common mode voltage level constant over the normal AGC voltage range. This compensation circuitry prevents the Output Amplifier stage from saturating. As the AGC voltage is reduced the common mode voltage level at R10 and R11 rise due to the reduced current supplied to Q1 and Q2. Also as the AGC voltage drops, the compensation circuitry increases the common mode voltage drop across R20 and R21 to keep the AGC amplifier output common mode voltage level constant.

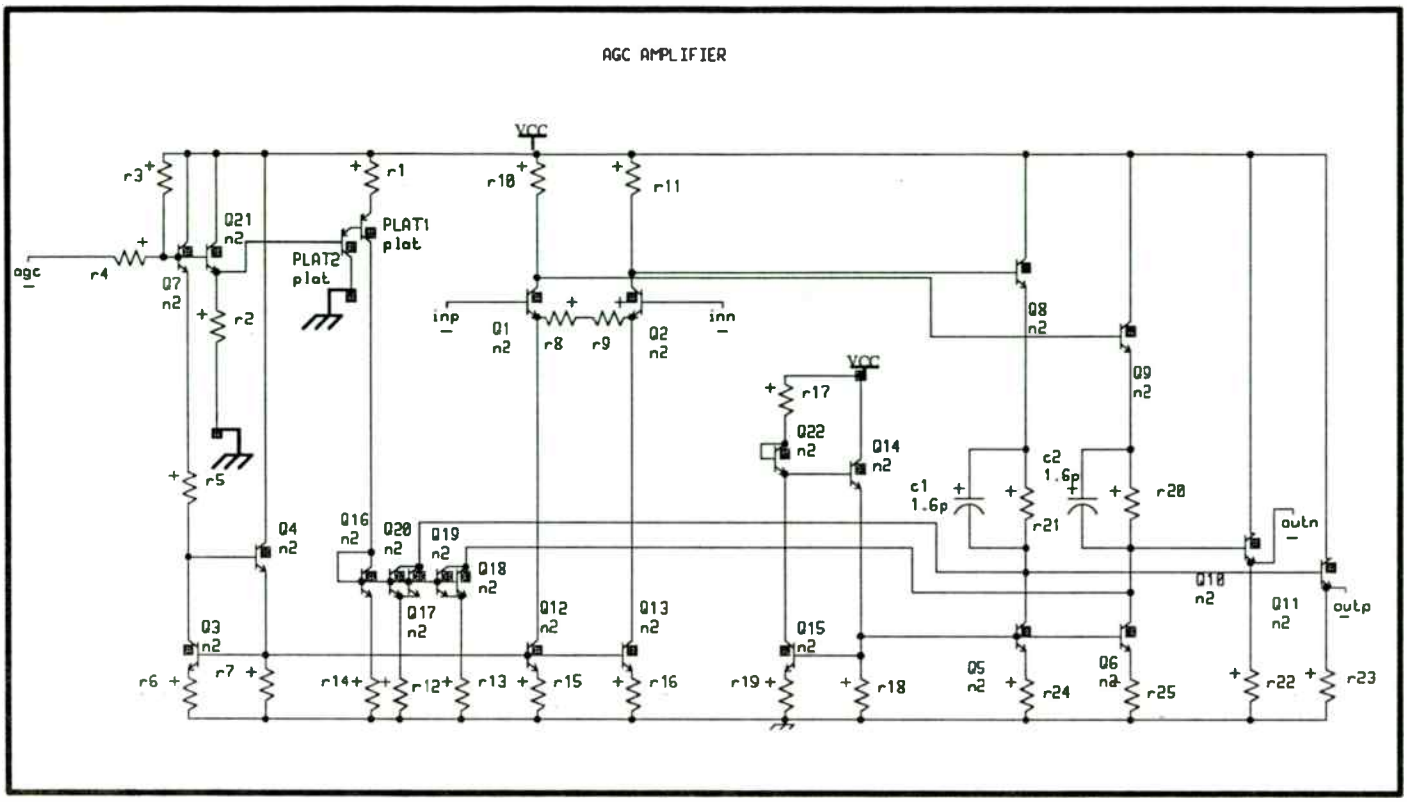


Figure 8. Simplified schematic of AGC amplifier

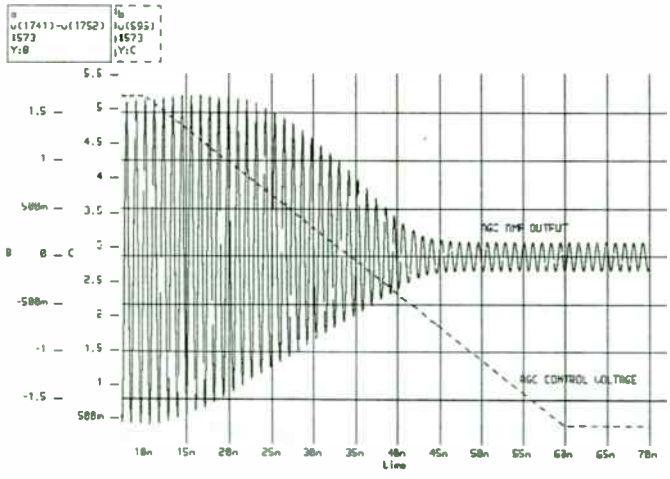


Figure 9. AGC effect on the transmitter output

OUTPUT DRIVER

The output driver initially performs a differential to single ended conversion with the amplifier composed of Q3 and Q18 (in Figure 10). The output of this stage is AC coupled through C11 to two emitter follower stages (Q7, Q12, Q24). These drive into the final common emitter stage consisting of Q21–23 and Q25. The output driver transfers power to a 37 ohm antenna through a matching network.

RESULTS

The initial results on the bipolar ASIC have been measured in isolation from the rest of the system to separate the performance of the ASIC from the total system. As shown in the plot of Figure 11, the measured output power was 18dBm which is 2 dB below our system goal of 20dBm. Since the tests were run in isolation from the rest of the system, the oscillator was not phase locked and thus the frequency is not exactly 915 MHz. Figure 12 shows a narrow band plot of the output signal as viewed on a spectrum analyzer. The resolution bandwidth of the measurement was 1 kHz. Extrapolating from this the phase noise measured 10kHz from the fundamental is 97.2 dB down for a 1 Hz bandwidth. Also specified was the match between the transmitter output power with the PN code in either state (logic 1 or 0). This output power with PN code in logic state 1 and in logic state 0 matched to within 0.5dBm. This result achieved the required level. All results were measured with a Tektronix 2756P spectrum analyzer.

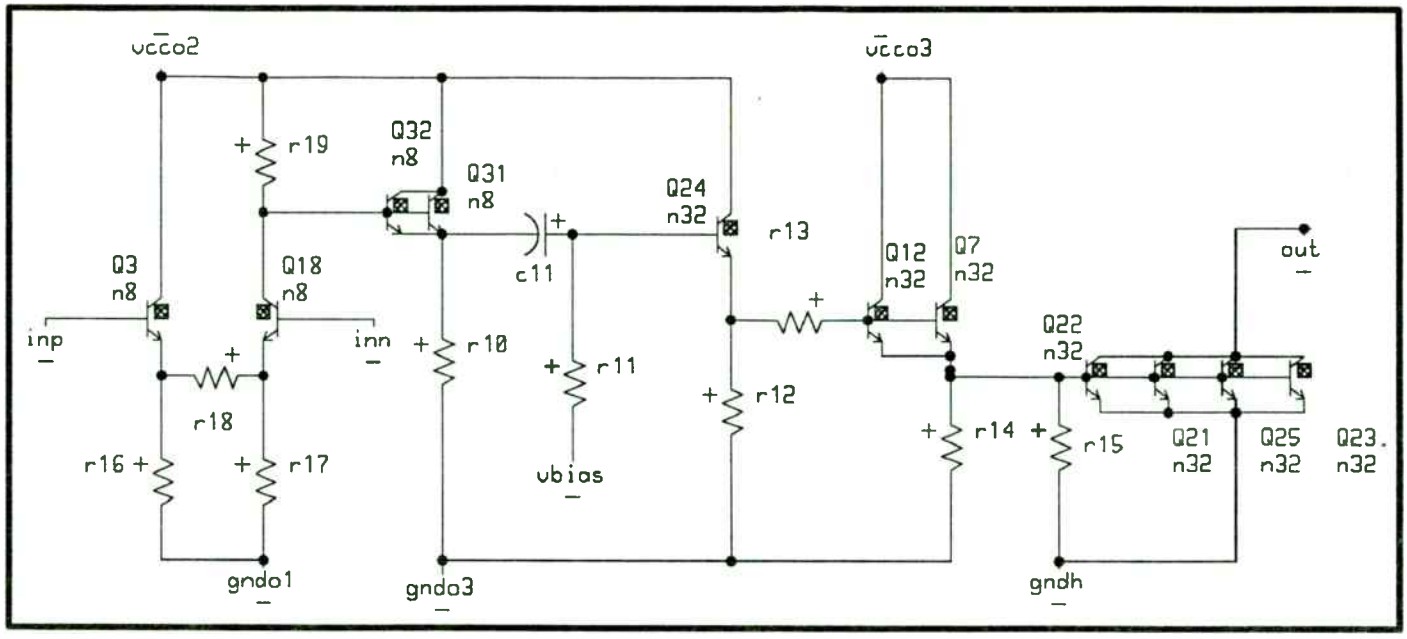


Figure 10. Simplified schematic of output driver

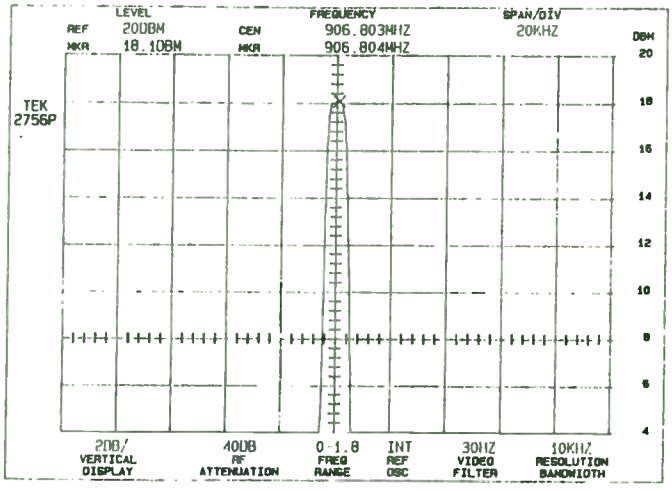


Figure 11. Transmitter output power

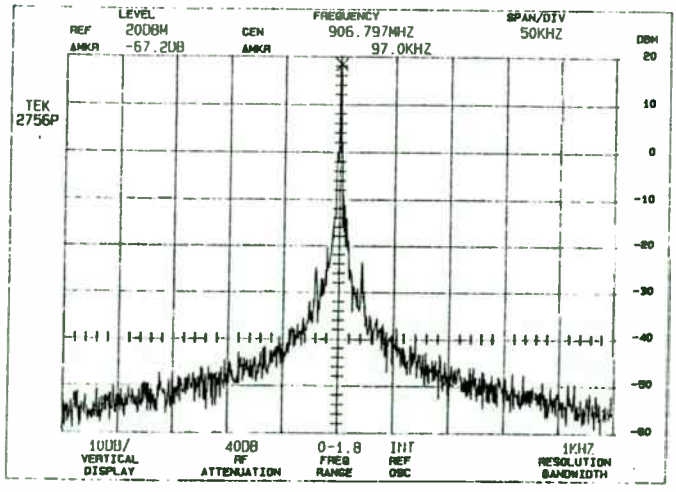


Figure 12. Phase noise at the output of the transmitter.

DISCUSSION

This design has successfully integrated five RF functions on a monolithic integrated circuit design. By so doing we have been able to make possible from a size and cost standpoint a system that would not have been possible with a discrete or multiple IC approach. With the rapidly growing use of the 915 MHz band this type of design will enable the implementation of many systems that would not otherwise be viable.

The RF ASIC met all of the system goals with the exception of transmitter output power. As discussed in the earlier section the measured output power was 2dB below the required 20dBm level. It is believed that the output stage is responsibile for this problem. There are a number of reasons indicating that the fault does not lie in earlier stages. In addition, there are several possible effects in the output stage which could cause the observed problem. These will not be discussed here. Further experimentation will need to be performed to test these theories.

REFERENCES

[1] "Quickchip 6 Integrated Circuit Design Guide", Version1, Tektronix, Inc., June 1989

[2] Gonzalez, Guillermo "Microwave Transistor Amplifiers, Analysis and Design", Prentice–Hall, Inc., Englewood Cliffs, N.J., 1984

[3] Hayward, W.H. "Introduction to Radio Frequency Design", Prentice–Hall, Inc., Englewood Cliffs, N.J., 1982

A Low Power RF ID Transponder

By Raymond Page
Wenzel Associates

This is the Grand Prize winner in the design category of the 1993 RF Design Awards Contest. This entry exhibited both innovative use of RF technology and an elegant implementation of that technology. The author was awarded a NOISE COM model UFX-BER noise generator for bit error rate testing.

For some time railroad companies have been wrestling with the problem of tracking rail cars. This has traditionally required manual log entry of identification numbers displayed on the cars as they pass through the switching yard. Some years ago, an effort was undertaken to use an optically scanned ID system. Dirt and optical registration problems led to its demise, forcing railroad companies to revert to the manual system. RF engineers have come up with a solution, using transponders mounted on the side of the cars which are read by interrogating transceivers positioned along the track.

Design Considerations

A practical transponder design must include minimal maintenance, a rugged low profile and low cost. The most elusive of these has been low cost. Presented here is a design which meets these requirements along with a brief discussion on the current state-of-the-art in passive RF identification transponders.

An important design constraint is that the transponder require little or no maintenance. Since no power is available from the rail car, the only conventional options are batteries or solar cells that maintain rechargeable batteries. The non-rechargeable batteries require periodic replacement, and the solar cell option would be both expensive and vulnerable to the environment. A passive design eliminates the need for batteries by rectifying energy from the interrogating RF field to power the circuitry.

The harsh environment presented to an RF device mounted on the side of a rail car is a challenging problem. Minimum clearance requirements, dirt, weather, vibration and an extremely large chunk of ferro-magnetic material near the antenna have to be considered. Additionally, the unit should be encapsulated. Microstrip patch antennas have come to the rescue. They afford a low profile and can be made with an ordinary double-sided printed circuit board. The patch antenna is on the top and a ground plane is on the bottom, thereby eliminating the effects of the steel mounting surface.

A Low Cost Transponder

An unusually simple method of converting the interrogating RF field into a data-modulated signal which can be transmitted back to the reader contributes to the low manufacturing cost of this transponder design. The circuit uses only one inexpensive microwave semiconductor (a diode) and allows all parts to be mounted on an FR-4 printed circuit board with the patch antennas (Figure 1). By contrast, other approaches use expensive microwave parts, including SAW devices, oscillators, mixers, filters and amplifiers. Designs involving more RF circuitry tend to be power hungry, requiring increased RF interrogation fields.

Figure 2 shows the block diagram of the low power transponder. A 915 MHz receive antenna powers the rectifier/frequency doubler/AM modulator. It provides a rectified DC source to the MCU which returns data to be AM modulated onto the doubled frequency. An 1830 MHz antenna transmits the modulated carrier.

A reader, incorporating an unmodulated 915 MHz interrogation transmitter with low (< –60 dBc) second harmonic distortion and an 1830 MHz AM receiver, is placed a relatively short distance away from where the transponder will pass (Figure 3). The amount of transmitted RF interrogation power needed to make the system function properly at a given distance can be estimated by equation 1:

$$P_r = P_t G_t G_r \lambda^2 / (4\pi R)^2 \qquad (1)$$

Where P_r is the received power, and P_t is the transmitted power radiated by an antenna of gain G_t. G_r is the gain of the receive antenna, λ is the free space wavelength and R is the distance between transmitters. G_t and G_r are the gains over an isotropic radiator. A sufficient second harmonic return path signal will occur for any combination of power gain and distance capable of energizing the MCU.

One watt of power transmitted with an antenna gain of 31.6 (16 dB) and received with an antenna gain of 2 (3 dB) allows the transponder to function from as far as 20 feet away. This suggests that just over 1 mW is adequate to energize the transponder.

The transponder's surprisingly low

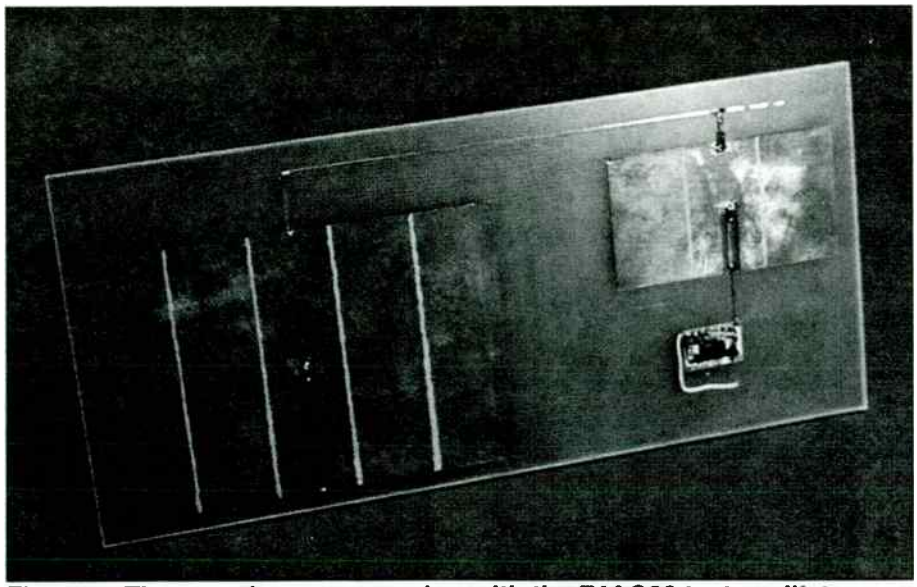


Figure 1. The complete transponder, with the 74AC00 test oscillator.

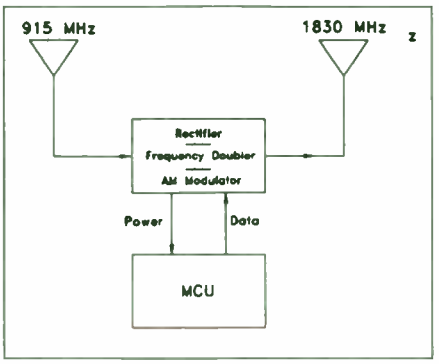


Figure 2. Block diagram of the passive transponder.

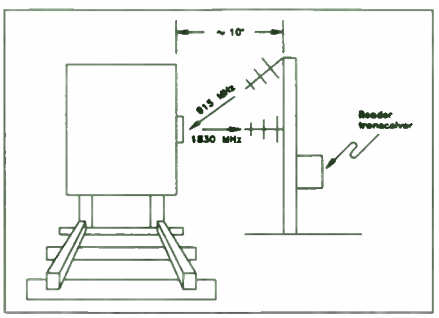


Figure 3. RF ID reader and transponder with rail car.

power requirement is due to its efficient means of rectification, frequency doubling and modulation. All of these functions are accomplished by a single microwave diode. A hybrid schematic in Figure 4 details the circuit. The 915 MHz patch antenna has two connections, a DC return path connected at the zero impedance point and a transmission line matched to the 120 ohm impedance at the edge of the antenna. The transmission line routes the signal to CR1 for rectification. A DC tap on the 1830 MHz antenna provides the power connection for the MCU. (See side bar on microstrip patch antennas.)

Careful placement of CR1 along the transmission line is crucial in creating the proper AC impedances for efficient frequency doubling. The 1830 MHz antenna becomes a 90 degree open stub at 915 MHz at the cathode of CR1, effectively giving the 915 MHz signal a low impedance trap to work against (Figure 5). Since the transmission line does not provide a similar low impedance on the anode side of CR1, a 90 degree open stub at 1830 MHz must be added.

Less than 100 uA are required to power the MCU (Figure 6). Consequently, little second harmonic is produced by CR1, leaving plenty of modulation headroom. Increased frequency multiplication occurs when the output port of the MCU goes low providing a path to ground for rectified current via the 1 kohm resistor, R1. Varying the value of R1 controls the modulation depth. CR2 and C2 work together to maintain sufficient voltage to the MCU while the voltage at C1 is being pulled down by the modulation action.

Performance

As previously noted, the system can operate up to 20 feet away. However, performance is measured at the 10-foot separation required during normal operation (Figure 7). For test purposes, a spectrum analyzer functions as the receiver. A 74AC00 gate oscillator in Figure 4 is substituted for the MCU to simulate load and logic level conditions. The oscillator simplifies confirmation of the concept. Three kHz modulation is used for easy detection by the analyzer.

The transponder transmits data at 94 percent AM modulation. Figure 8 is the detected 3 kHz square wave. Measurements of the rectified voltage (2.7 VDC) and current (1.45 mA DC) give 3.9 mW total power which correlates nicely with the received power (5.3 mW) predicted

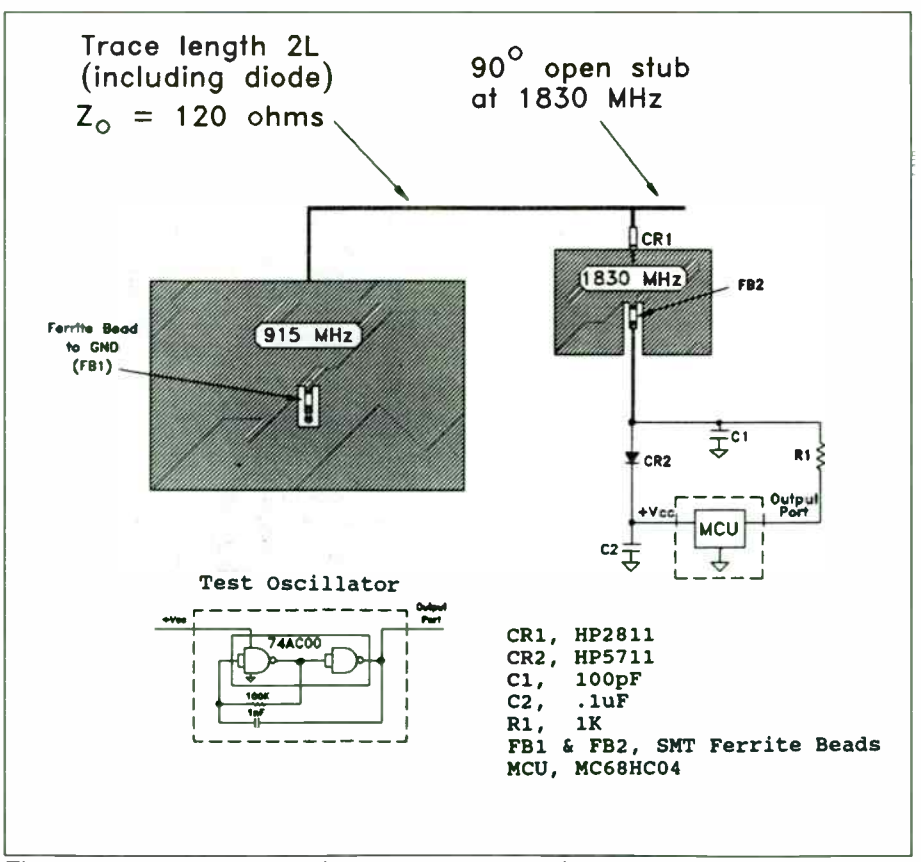


Figure 4. Hybrid schematic of transponder circuit.

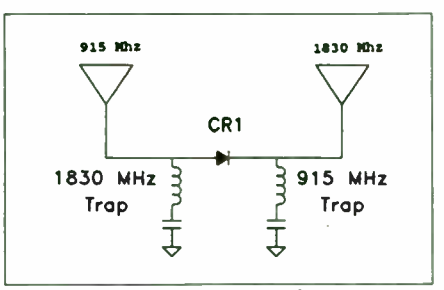


Figure 5. Equivalent AC circuit of transponder showing RF traps.

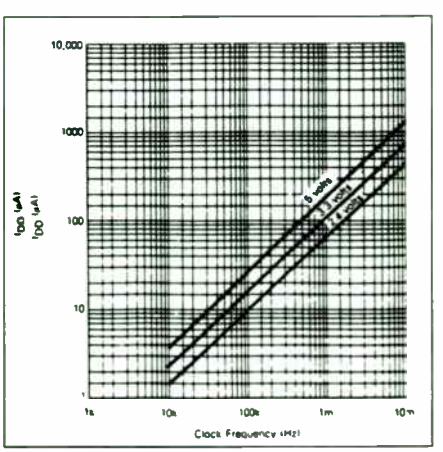


Figure 6. Current vs. clock frequency for a typical 68HC04 MCU.

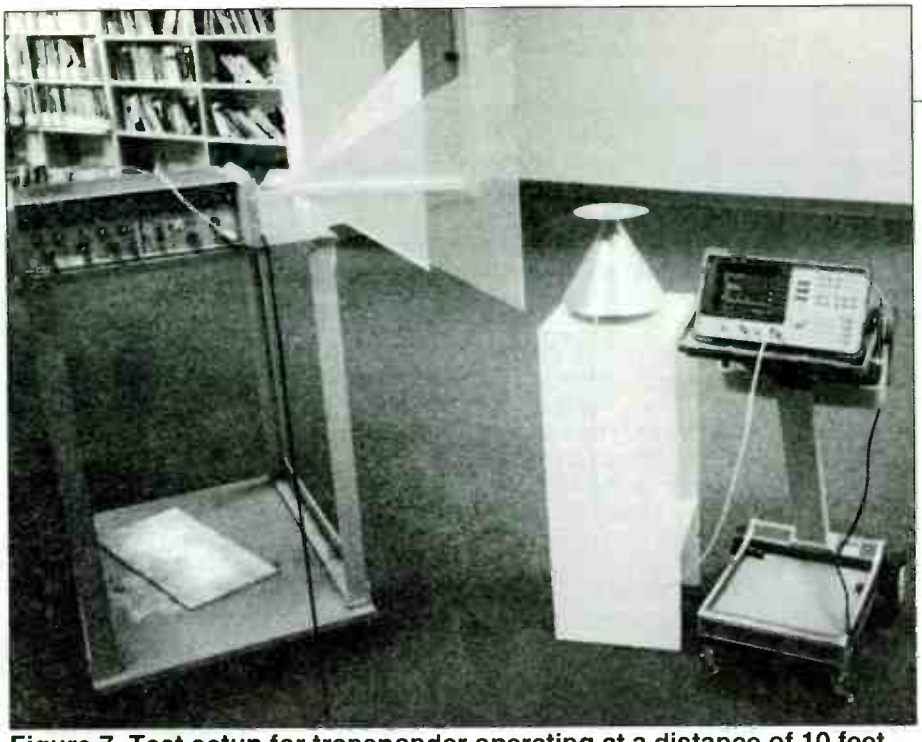


Figure 7. Test setup for transponder operating at a distance of 10 feet.

by equation 1 at a distance of 10 feet.

Improvements

Inherent compatibility with spread spectrum is provided by this design since the returned signal frequency is derived directly from the interrogation signal. Frequency spreading is limited only by the bandwidth of the patch antennas. With the simple addition of a micro-power line receiver and the associated communications software, the transponder can be upgraded for two-way information applications. Size reduction can be accomplished by increasing operating frequency at the expense of costlier substrate material. Borrowing technology from missile and aircraft radar technology the transponder could be made a part of the "skin" of its host.

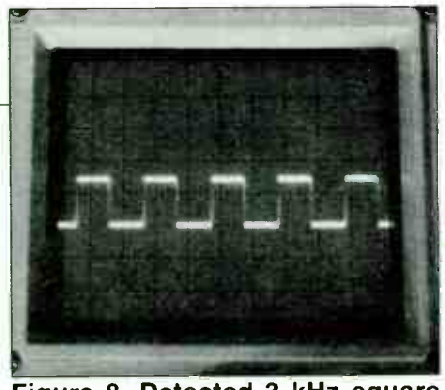


Figure 8. Detected 3 kHz square wave.

Summary

This paper has described the design, operation and application of a low-power RF identification transponder. The simple design is spectrum friendly, requiring

Rectangular Microstrip Patch Antenna

The rectangular patch antenna is essentially a resonant microstrip with an electrical length of 1/2 the wavelength of the frequency to be transmitted or received. Microstrip patch antennas work well for applications requiring a low profile, offering a height equal to the thickness of the printed circuit board from which they are made. PTFE substrates are normally used to minimize dielectric losses which affect the efficiency of patch antennas. However, FR-4 is a cost effective alternative for low power applications at frequencies below 2 GHz.

Microstrip antennas come in all sizes and shapes. A rectangular patch is chosen for its simple geometry and linear polarization when fed from the center of an edge. The input impedance varies as a function of feed location. The edge of a 1/2 wavelength antenna has an input impedance of approximately 120 ohms which drops to zero ohms as the feed point is moved inboard to the center of the antenna. This allows easy impedance matching and provides a convenient means of DC tapping the antenna as seen in the transponder design.

For simplicity, the dimensions of the microstrip patch antennas in Figure 9 are in terms of L, which is equal to 1/2 the electrical wavelength of the receive antenna (915 MHz). L can be determined by equation 2:

$$L = 0.49(\lambda/\varepsilon_R) \qquad (2)$$

where λ is the free-space wavelength and ε_R is the relative permittivity of the printed circuit board.

Bandwidth is determined by the substrate thickness and can be approximated for an SWR of less than 2 by equation 3:

$$BW = 128f^2t \qquad (3)$$

BW is in MHz, f is the operating frequency in GHz, and t is the substrate thickness in inches.

Applying equations 1 and 2 to the transponder design using 0.125 inch FR-4 substrate material with an effective permittivity of 4.7 results in a value of 2.92 inches for L and a bandwidth of 13.4 MHz at 915 MHz.

Our Design Contest Winner

Raymond Page is a design engineer for Wenzel Associates, a manufacturer of high performance crystal oscillators and frequency standards. Ray specializes in low noise designs for devices including oscillators,

phase locked frequency sources, multipliers and dividers. In addition to having fun with electronics, he enjoys outdoor sports and music. He can be reached at Wenzel Associates, 1005 La Posada Drive, Austin, Texas 78752, or by telephone at (512) 450-1400.

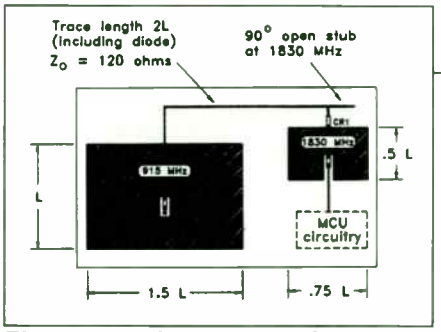


Figure 9. Dimensions for patch antennas.

minimal interrogation power and allows easy conversion to spread spectrum without modification to the transponder. Designed with one inexpensive microwave part on a single piece of FR-4 substrate, component and manufacturing costs are kept down, potentially opening up markets served exclusively by bar coding technology. Other uses include automatic tolling, inventory tracking and military vehicle security. *RF*

References

1. Howard W. Sams & Co., *Reference Data for Radio Engineers*, Chapter 27, Sixth Edition, 1977.
2. Robert E. Munson, "Conformal Microstrip Antennas," *Microwave Journal*, March 1988, pp. 91-109.
3. Alan Tam, "Principles of Microstrip Design," *RF Design*, June 1988, pp. 29-34.

MULTI-COMPONENT MODULE FOR HIGH SPEED PASSIVE DESIGN

Mark Brooks
Thin Film Technology
1980 Commerce Drive
N. Mankato, MN 56003

ABSTRACT

As electronics becomes evermore sophisticated in terms of advanced semiconductor devices, there will always remain a need for passive components to provide the necessary support functions. Resistors, capacitors, inductors, and transmission line components exist as a passive subset to high performance designs. Notwithstanding, advancements in the packaging of such devices must go hand in hand with the evolution of other circuit counterparts in order to realize total system improvement. In order to satisfy these requirements, a technique to allow the assembly of high performance thin film devices on ceramic into a sophisticated multilayer package has been devised, with circuit topologies limited only be the designers imagination.

Scope

This paper will outline the approach of Thin Film Technology Corporation for high speed passive component design utilizing multilayer ceramic. Construction of a high speed delay line is first presented, followed by a discussion for application of the technology to produce a Butterworth filter.

Passive Multilayer Component Module Introduction

Though not an offspring of Multi-Chip Module (MCM) technology, this Passive Multilayer Component Module (PMCM) technology parallels the MCM approach; but shines in it's processing simplicity as opposed to conventional MCM types and applications:

MCM-L: laminated board approach
MCM-C: cofired ceramic approach
MCM-D: tradional thin film aproach
PMCM: thin film on via processed sintered ceramic

These PMCMs offer an approach that allows for substrates containing vias to be metallized by thin film processing and provide for electrical interconnect. A novel technique to select appropriate alloys and alloy thicknesses, however, is also cause for a mechanical connection between layers. By depositing different alloys at the surface sites of two adjacent vias to be connected, the prerequisite for combination into a binary allow is created. After micron level alignment of these metallized vias to one another, sufficient heat to induce a liquidus state to the metals is provided. Subsequent diffusion of the different metals results in a different melting point alloy, one that is appreciably different than that of the initial metals. This melting point is higher than that achieved in conventional circuit board assembly production, and consequently serves to mechanically hold the interconnect intact.

The PCMC's show improvement for a variety of designs and offer advantages to other technologies by realizing flexibility

alumina reduces dependency on cofired ceramic suppliers, as vias can easily be machined in house. The ability to produce the PMCM with sintered ceramic reduces tooling costs, eliminates via location tolerance drift, and provides custom via patterns as needed substrate by substrate. Standardized fabrication of subassembly components allows for combination of multiple circuits elements into custom circuits with ease. Tape and reel packaging of individual chip subassemblies provides for pick and place assembly in a highly automated production line. Standardized input/output cover ceramics provides final form factor variations so that the same module may be packaged either as a leaded device, face down solder bump, or for wire bonding.

Furthur added benefits include high speed circuit applications, where microwave circuitry can perform due to the absence of leads and lead length for the interconnect. This dramatically reduces inductance, which is critical for high speed designs. Solder bump terminations of the PMCM vehicle realizes the shortest length possible from the circuit board trace to the device elements. Coplanar stripline termination pads furthur improve high frequency performance providing impedance control. With this recontruction of multilayer ceramic technology, a viable method is available for commercial microwave applications.

CONSTRUCTION

A typical thin film sequence is used for the fabrication of the PMCM device. The fabrication of each via ia accomplished by ceramic machining, with appropriate control of program and process parameters to acheive the necessary via quality.

In the first step, a sintered ceramic substrate (96% to 99% alumina oxide) is positioned to a holder, where computer control of X, Y, and Z axes is maintained in relation to the maching tool. At each via location, table movement of the Z axis may raise and lower the substrate coincident to the machining. Each via is precisely formed into the substrate. Because the substrate is machined in a sintered state, concern for the shrinkage movement that occurs when vias are processed in green state cermic is eliminated.

Substrate thickness and composition are selected for effective via processing and cost consciousness. An important additional benefit of the via is to use a grid or nest of vias for thermal management of a particular area. Metallized blind or through vias can be fabricated to serve this purpose.

Thin film processing techniques applied in the following steps later define the metallization and features of the circuit patterns, which may include combinations of resistors, capacitors, inductors, and transmission line structures. The delay line device described contains microstrip transmission line on individual ceramic substrates, which will become of stripline design as the ceramics are stacked adjacent to one another. Of particular importance is the plating processing step, where the appropriate thicknesses of three metals is achieved. These metals fill the via conductor hole, and are then plated to a "bump" on the substrate surface. Plating thickness control is responsible for the different metals final atomic weight % ratio, which is necessary to allow for the correct binary alloy result during the assembly lamination process.

Dicing or laser scribing and breaking each substrate into individual pieces allows for tape and reel packaging of separate subassemblies. This prepares the product for pick and place processing into the correct "stack" sequence for final assembly (Figure 1).

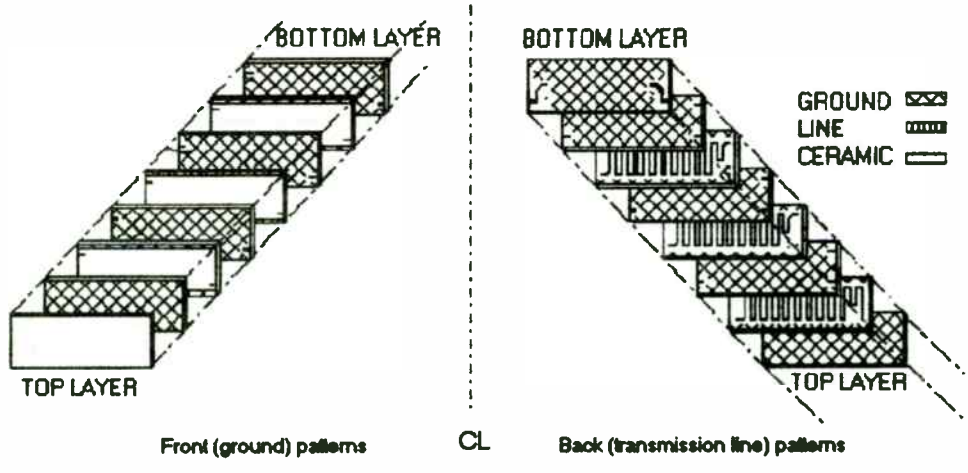


Figure 1

Assembly of the product follows in a clean room envrionment. Alignment of the subassemblies into lamination tooling is prerequisite for successful lamination to follow. With several "stacks" loaded into a lamination jig, proper pressure, temperature and time, and atmosphere is provided to impart a liquidus state to the plated bumps adjacent to one another. The metals then diffuse into one another, and create a substrate to substrate bond after cooling.

The final processes serve to package the product into a surface mount configuration and verify AC and DC performance. The gullwing form factor (Figure 2) follows traditional SMT packaging. Alternate packaging schemes include face down solder bump (flip chip) and a non-molded approach for wire bonding.

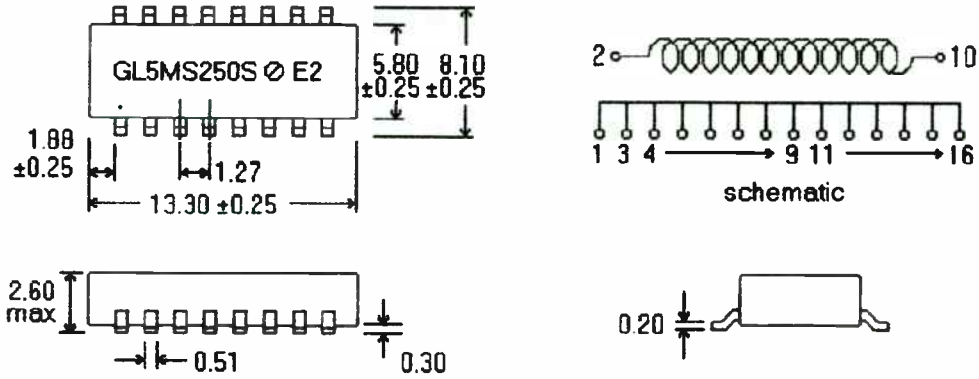
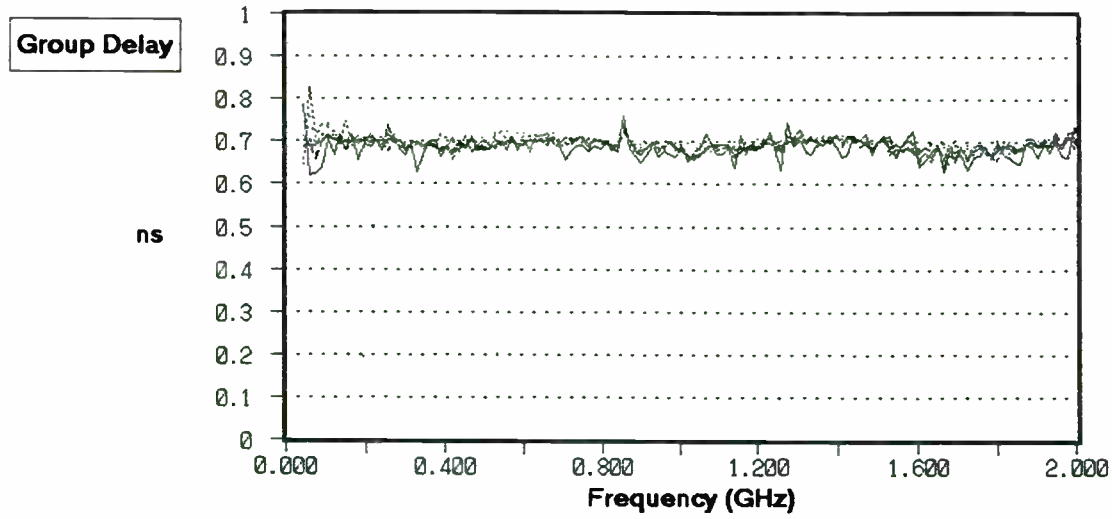
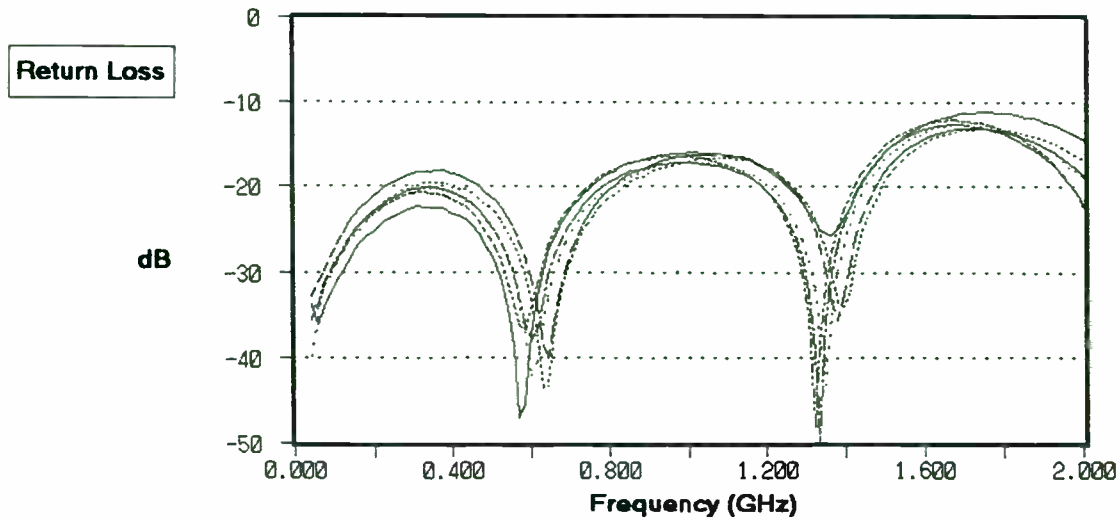


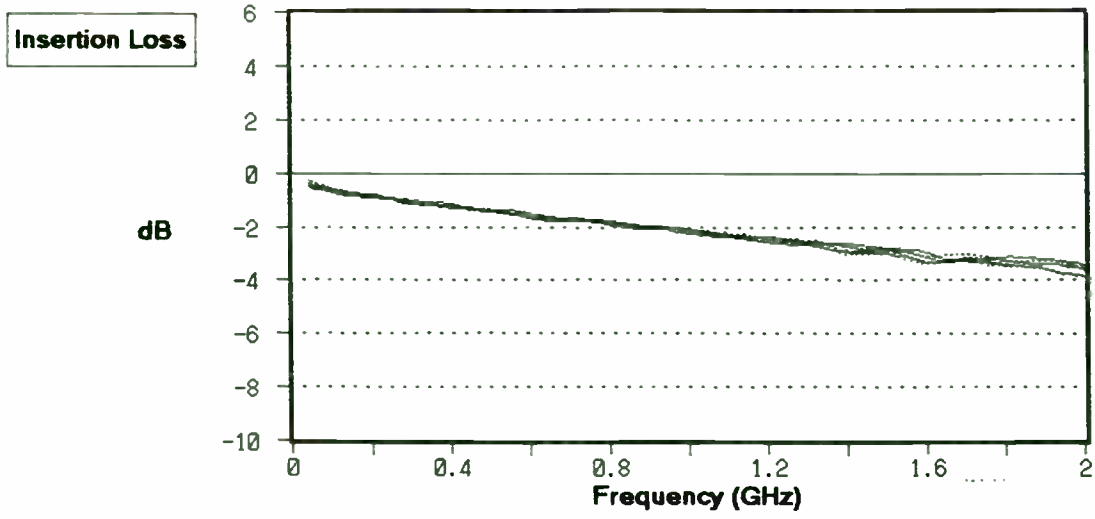
Figure 2

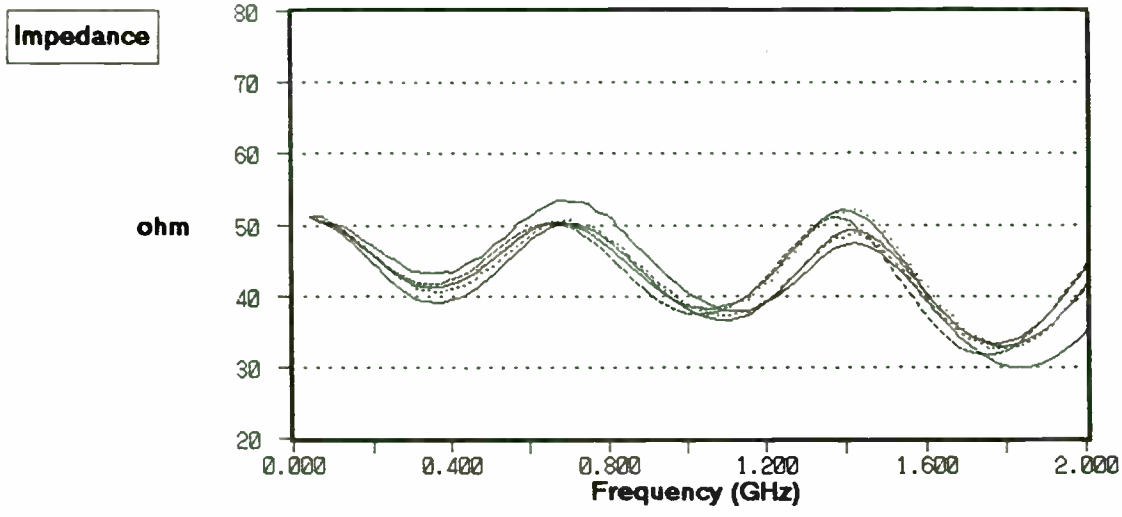
PERFORMANCE

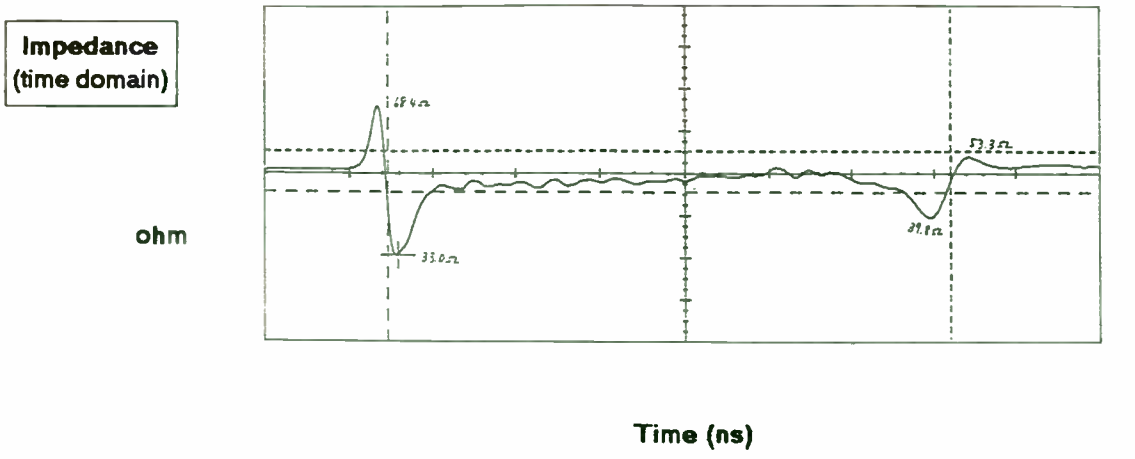
Performance for a 0.5 nanosecond device can be seen as follows. Five devices were tested for group delay, return loss, insertion loss, and impedance.











Time delay is varied by more or less microstrip and ground substrates in each stack. The above 0.5 ns device is comprised of four ceramic layers. Twelve layered stacks are common. Additional time delays make the component taller, and affect system and board design vertically, but not in terms of square surface displacement. Top and bottom ground planes provide shielding for interference.

DIRECTION

A simple Butterworth filter is made possible by the PMCM approach. Using spiral inductors with vias, and capacitors utilizing the ceramic as the dielectric, this design provides superior performance to discrete mounted SMT components. Additional products, such as sophisticated high density resistor or resistor/capacitor networks are also possible. Using multiple layers in parallel allows a current sense resistor (<2 ohms) to be fabricated. Many commercial RF products can utilize PMCM technology for cost competitive applications.

For feedback or questions please contact Mark Brooks. Correspondence with the author is offered Wed. at 507 625-8445.

DESIGN OF A SEARCH BASED PLL

Michael F. Black
MS 3145, 8505 Forest Lane,
Texas Instruments
Dallas, TX 75266

Before a phase lock loop can lock some mechanism must move the VCO and the input signal close enough for acquisition. Different frequency ranges and circuit requirements will allow different circuit topologies to achieve proximity. Logic level loops have the advantage of simple flip flop arrangements that combine a frequency discriminator detector with a $\pm 2\pi$ linear phase detector. Self correcting detectors with frequency discriminators become much more complicated when the PLL input is a wide dynamic range analog signal or a pulsed signal. Search circuits become a simple alternative for lock acquisition with these types of phase lock loops.

Search circuits can provide either an open or closed loop action. The search circuit should sweep the VCO control voltage until lock is acquired then shut down. The lock action will normally be strong enough to overcome the search action. A closed loop sweep will sense lock and disable itself. Sweep is reinitiated with loss of lock. Gardner[1] has described a unique circuit of this type as illustrated in Figure 1. The opamp loop filter serves also as a low frequency sine wave oscillator. The feedback network from the output to the non inverting input sets the oscillator frequency. This network could be a Wien bridge or a shunt T network. Before PLL closure the oscillator output voltage sweeps the VCO. When the VCO frequency moves across the input, the phase detector provides an error voltage which slows the sweep rate and drives the loop into lock. The feedback with the phase lock loop closed is enough to overcome the positive feedback of the loop filter oscillator circuit and it stops oscillating. As simple and interesting as this circuit may seem, its applications are limited. The sine wave sweep is inefficient and the circuit elements are very critical. The loop filter design goals may have to be compromised to satisfy the additional oscillator requirements.

Figure 2 illustrates another approach to frequency sweep and search. The loop filter again does double duty. During search it acts as an integrator. A constant current input provides a voltage ramp to the VCO. With lock the current source turns off. The loop filter closes the loop and controls the VCO.

A circuit topology of Figure 2 provides a type two second order loop. This is one of the most versatile and forgiving PLL circuits for analog input requirements. Additional filtering or phase margin optimization can be gained by changing the loop filter to the third order circuit as shown in Figure 3, but for this paper the circuit of Figure 2 will be assumed.

With an ideal opamp, the PLL transfer function of the loop in Figure 2 can be described by the well documented second order equation. The parameters damping ratio and natural frequency describe the time domain and frequency domain response. The loop design parameters include the VCO gain, the phase detector gain, the frequency divide ratio, the loop bandwidth and the damping ratio. The VCO gain, Kvco, is the measured tuning slope in Hz/volt. With an analog phase detector such as a mixer the typical error voltage output is a cosine wave where the amplitude is proportional to the phase difference between the two inputs and the amplitude of the input signal. For a given signal level the phase detector gain, Kpd, in Volts/Rad is the slope of the cosine wave at the zero crossing. The slope is equal to the zero to peak output voltage. Damping ratio, δ ,and natural frequency, fn, are design goals set by the system requirements. Design nomograms illustrate the performance expected with normalized bandwidths at various damping ratios. Figures 4, 5, and 6 plot the relationship between the 3 dB bandwidth, the 0 dB bandwidth, the phase margin and selected values of damping ratio and natural frequency.

The loop filter determines the closed loop operation of the PLL. The standard design methodology will determine the required filter values based solely on this closed loop criteria. For a typical PLL, the VCO gain, Kvco, and the phase detector gain, Kpd, will be known values. The frequency divide ratio, N, is set by system requirements. A value of N = 1 will be used. The damping ratio, δ , and the natural frequency fn or 3 dB bandwidth are the design goals. The loop filter has three unknown components, Rinput, Rfb and Cfb . With two design goals and three unknowns an infinite number of solutions results. So pick a value for Cfb and then determine Rfb and Rinput.

$$(1)\ R_{fb} = \frac{\zeta}{\pi * f_n * C_{fb}}$$

$$(2)\ R_{input} = \frac{K_{vco} * K_{pd}}{\pi * f_n^2 * C_{fb} * N}$$

The design nomogram in Figure 4 describes the relationship between natural frequency and 3 dB bandwidth as a function of damping ratio. As an example assume a loop with a VCO that tunes at 100 MHz/volt and a phase detector that produces ± 190 mV output. For this example the system analysis indicates the damping ratio should be 2.0 and the 3 dB bandwidth 300 KHz. For these values Figure 4 shows that the natural frequency must be 71 KHz. Figures 5 and 6 indicate the 0 dB open loop bandwidth will be 284 KHz and the phase margin will be 86 degrees.

The feedback capacitor is picked to be .1 ufd. Then from equations 1 and 2, Rfb = 89.9 and Rinput = 6033. Figure 7 illustrates this configuration.

As discussed earlier this circuit has no means of signal acquisition. It will only capture and lock if the VCO and input frequency are close. Figure 8 illustrates modifications that will implement a sweep function using the loop components. The switched current sources will cause the loop filter to act as an integrator before lock and produce a positive and negative slope linear waveform. Voltage comparators limit the sweep range to levels slightly greater than required to allow the VCO to cover the desired frequency range. The basic operation is simple and the hardware is simple but the design parameters must be carefully controlled to insure initial lock and the correct closed loop performance.

During the search phase one of the current sources injects current into the loop filter opamp inverting input. The opamp is in a stable closed loop mode during search but the PLL is not locked and the loop is not closed. As the VCO sweeps across the input frequency an error voltage develops at the phase detector output. This is a cosine wave equal in frequency to the difference between the input and the VCO frequencies. When this error frequency comes within the PLL loop bandwidth, feedback will push the VCO closer to the input and lock will occur. With lock the VCO must stop sweeping. This implies that the VCO control voltage is now constant. However search current is still being injected. The rules of operation for a closed loop opamp require that the net current into the inverting node be zero. Since this search current is not being used to sweep the output voltage it must go somewhere. It is in fact sunk out from the inverting input through Rinput. If the search current is I then a voltage equal to I x Rinput will be measured at the phase detector output while the search current is on . If it is turned off the phase detector output will drop to zero to match the non inverting input voltage. However it is easy to pick a combination of values that create a no lock circuit. Consider the values used in the example circuit of Figure 8. If the search current were chosen to be .1 mA, at the instant of lock a voltage of Rinput x I = 603 mV must be present at the phase detector output. But from the initial design values, the phase detector is only capable of producing 190 mV. As the VCO sweeps across the input frequency, lock will not occur. A small disturbance will be observed in the VCO sweep voltage at crossover, but the sweep will not stop. So a set of design goals and equations are needed to insure both lock and correct closed loop operation after lock is achieved.

Figure 9 illustrates the design parameters for a versatile search derived loop. Several items must be determined by the system requirements and external interfaces.

The standard PLL loop parameters of damping ratio and bandwidth must still be part of any design set. Next is the sweep time, Ttotal. This example allows for a different up and down time, Tup and Tdn. The characteristics of the VCO will determine the upper and lower sweep voltage, Vvcoup and Vvcodn, required. The values chosen should allow margin for tolerances to guarantee a useable lock range. The last system level design parameter is the up and down phase detector voltage levels, Vpdup and Vpddn. These are the voltage levels at the phase detector output at the moment of lock when the up or down search current is on. These voltages will be opposite in sign because of the reversal of the search current for up and down sweep. These two levels need to be carefully chosen. Correctly used they can provide a valuable amplitude discrimination action to limit the range of acceptable signals. Figures 10 and 11 illustrate typical transfer curves for a phase detector. Note that the positive and negative values are not normally the same. These graphs plot the phase detector peak voltage outputs vs the amplitude in dBm of the input. For example the system requirements might dictate that the loop should first try for lock to inputs greater than -2 dBm. Failing to lock at greater than -2 dBm the loop could then try for greater than -6 dBm. From the graph in Figure 10, -2 dBm will produce a 175 mV output and from Figure 11, -6 dBm will produce a -100 mV output. If the value of sweep current I and the loop filter input resistor Rinput are chosen with these levels in mind then the loop can be constructed to insure no lock below these levels. These two levels define the lock threshold parameters Vpdup and Vpddn.

The system level design parameters include damping ratio, natural frequency, frequency division ratio, Tup, Tdn, Vvcoup, Vvcodn, Vpdup and Vpddn. The design constants are the Vco gain, Kvco, and the phase detector gain, Kpd. The design equations must determine Rinput, Rfb, Cfb , Iup, and Idn. to insure both correct search and lock. Equations 3, 4 and 5 develop the ratio between Iup and Idn.

$$(3)\ V_{pdup} = I_{up} * R_{input}$$

$$(4)\ V_{pddn} = I_{dn} * R_{input}$$

$$(5)\ \therefore I_{dn} = \frac{V_{pddn}}{V_{pdup}} * I_{up}$$

The ramp up time and the ramp down time can be found from the equations for a constant current source driving an integrator.

$$(6)\ V_{vcoup} = \frac{I_{up} * T_{up}}{C_{fb}}$$

$$(7)\ V_{vcodn} = \frac{I_{dn} * T_{dn}}{C_{fb}}$$

$$(8)\ \text{LET } A = \left(\frac{V_{pddn}}{V_{pdup}}\right) * \left(\frac{V_{vcoup}}{V_{vcodn}}\right)$$

$$(9)\ \text{THEN } T_{up} = A * T_{dn}$$

The value of the up and down current sources and the loop filter input resistor come from rearranging equations 6 and 7 and then solving for Rinput.

$$(10)\ I_{up} = \frac{V_{vcoup} * C_{fb}}{T_{dn}}$$

$$(11)\ I_{dn} = \frac{V_{vcodn} * C_{fb}}{T_{dn}}$$

$$(12) R_{input} = \frac{V_{pdup}}{I_{up}}$$

$$(13) \therefore R_{input} = \frac{V_{uppd} * T_{up}}{V_{vcoup} * C_{fb}}$$

The final value required for the loop filter is the feedback resistor, Rfb. In the earlier example the three unknowns, Rfb, Rinput and Cfb had a infinite set of solutions with only the two design goals of damping and natural frequency. These are still two of the design goals but the search constraints have reduced the size of the solution set. Cfb was picked early in the search equation derivation. Rinput was derived from the threshold and sweep range values. With only Rfb left either the damping ratio must be chosen and the resultant natural frequency calculated or vice versa. Picking the damping ratio is usually the most practical choice.

To find Rfb and the natural frequency equations 1 and 2 may be rearranged.

For a PLL with a damping ratio ς, a VCO gain Kvco, a phase detector gain Kpd, and a divide ratio N

$$(14)\ R_{fb} = \zeta * \sqrt{\frac{2 * N * R_{input}}{\pi * K_{vco} * K_{pd} * C_{fb}}}$$

$$(15)\ f_n = \sqrt{\frac{K_{vco} * K_{pd}}{2 * \pi * N * C_{fb} * R_{input}}}$$

Applying these equations to the original example will result in a fully defined search driven PLL.

For Kvco = 100 MHz/volt
Kpd = .190 V/RAD
Vpdup = .175 volt
Vpddn = .100 volt
Vvcoup = 7
Vvcodn = 1
Total search time = 50 ms
and again picking Cfb = .1uf

Equations 10 through 13 yield

Iup = 33.0mA
Idn = 18.9mA
Tup = 18.2mS
Tdn = 31.8ms
Rinput = 5303

Repeated iterations for values of Rfb for various damping ratios shows the effect on bandwidth and damping as listed in Table 1. Although none of the combinations of 3 dB bandwidth and damping match the design goals exactly the closest match comes with Rfb = 81.

Figure 12 illustrates the schematic of the final design. Circuit details of the window comparators and logic controls have been omitted for clarity. Figure 13 shows the loop bandwidth and Figure 14 plots the sweep voltage when the circuit is allowed to free run. Design verification of the performance comes from both SPICE simulation and actual hardware measurements.

Shutting off the search circuit is another topic that is very dependent on system requirements. The easiest circuit is shown without detail in Figure 15. This uses a retriggerable one shot that is fired each time the sweep voltage trips the high or low level of the window comparator. The one shot time out is set longer than either the up sweep or down sweep time. If the one shot is not retriggered before it times out, then the assumption is made that the loop has locked and the VCO voltage is at a constant value. When the one shot times out, the search current is turned off and the phase detector offset voltage returns to ground.

More demanding systems may require full lock verification circuits with signal injection or quadrature detectors. With these circuits only a full lock confidence check will shut down search.

CONCLUSIONS

The conflicting PLL requirements of efficient search for lock and loop bandwidth after lock can be resolved with a versatile set of equations and design methodology. With a defined set of system requirements and component parameters a simple loop filter/sweep generator can be designed. The measured results can be expected to closely match the predicted performance.

This search system matches the component values to both the lock and search requirements. Search times and lock thresholds determine the sweep currents and the natural frequency. The selection of a feedback resistor sets the closed loop 3 dB bandwidth, the phase margin and the transient response.

References

1. F.M. Gardner, *Phaselock Techniques*, John Wiley & Sons, New York, 1979, p82

FEEDBACK RESISTOR	DAMPING RATIO	ZERO DB BW Hz	PHASE MARGIN	THREE dB BW Hz
75.0	1.779	2.69547E+05	85.497	2.89278E+05
76.0	1.803	2.73096E+05	85.613	2.92600E+05
77.0	1.827	2.76649E+05	85.726	2.95793E+05
78.0	1.850	2.80203E+05	85.833	2.99113E+05
79.0	1.874	2.83758E+05	85.937	3.02439E+05
80.0	1.898	2.87317E+05	86.037	3.05798E+05
81.0	1.922	2.90874E+05	86.134	3.09140E+05
82.0	1.945	2.94433E+05	86.227	3.12430E+05
83.0	1.969	2.97992E+05	86.316	3.15889E+05
84.0	1.993	3.01551E+05	86.403	3.19231E+05
85.0	2.016	3.05114E+05	86.487	3.22574E+05
86.0	2.040	3.08680E+05	86.567	3.25829E+05
87.0	2.064	3.12243E+05	86.645	3.29221E+05
88.0	2.088	3.15805E+05	86.721	3.32643E+05
89.0	2.111	3.19375E+05	86.793	3.35950E+05
90.0	2.135	3.22939E+05	86.864	3.39427E+05

Table 1
Loop Parameters as a Function of Loop Filter Feedback Resistor

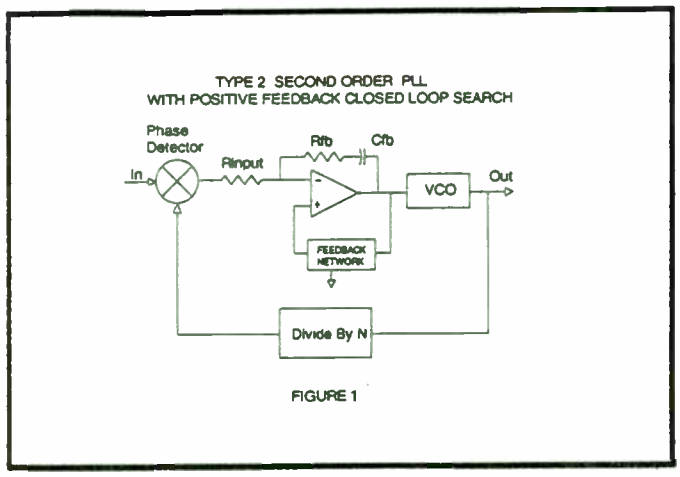


FIGURE 1

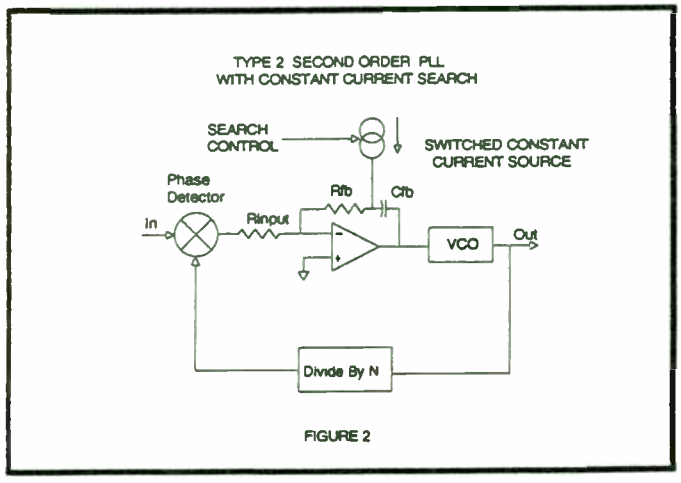


FIGURE 2

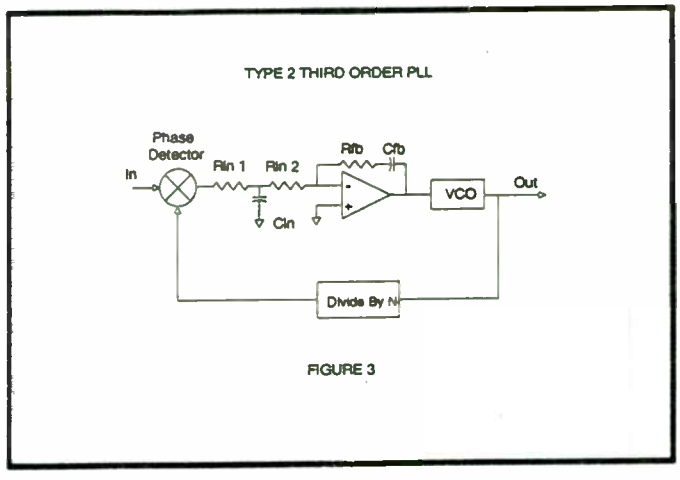


FIGURE 3

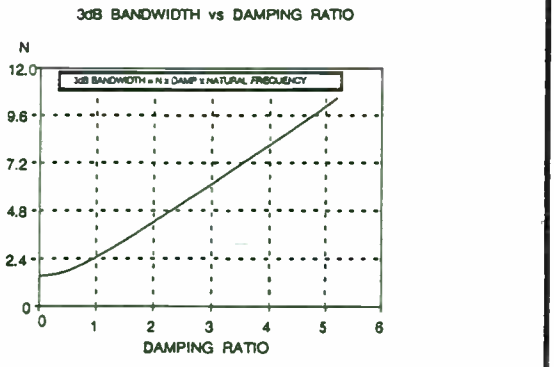


FIGURE 4

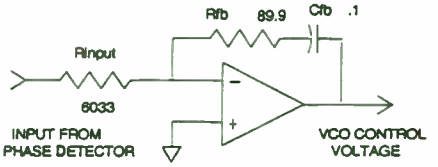


FIGURE 7

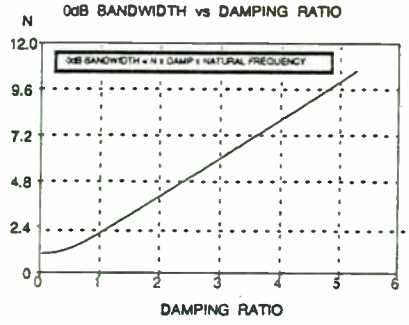


FIGURE 5

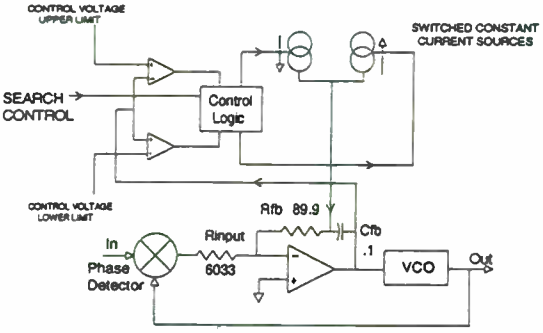


FIGURE 8

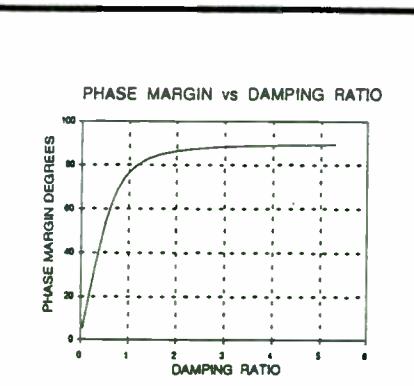


FIGURE 6

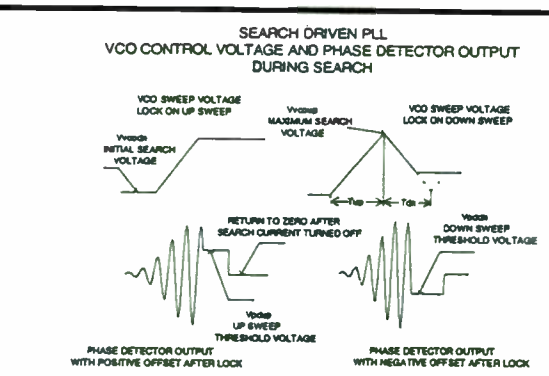


FIGURE 9

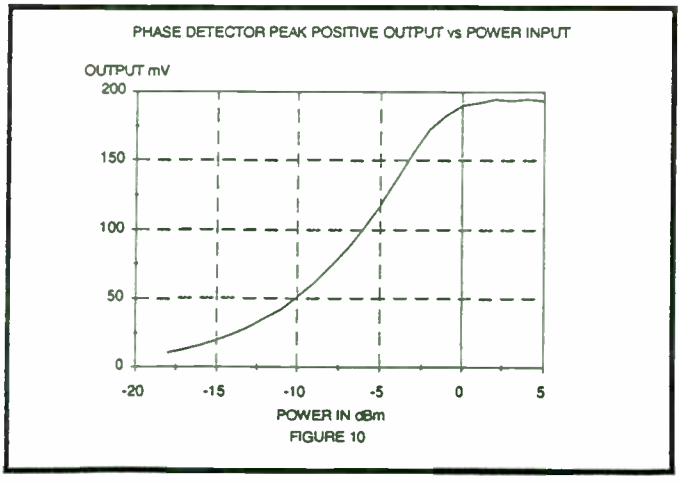


FIGURE 10

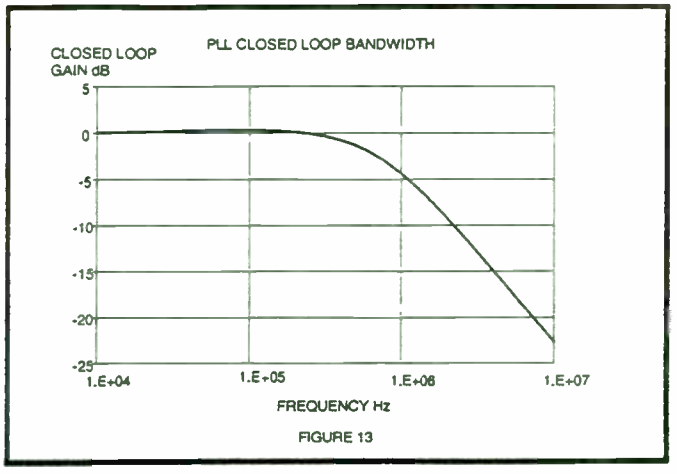


FIGURE 13

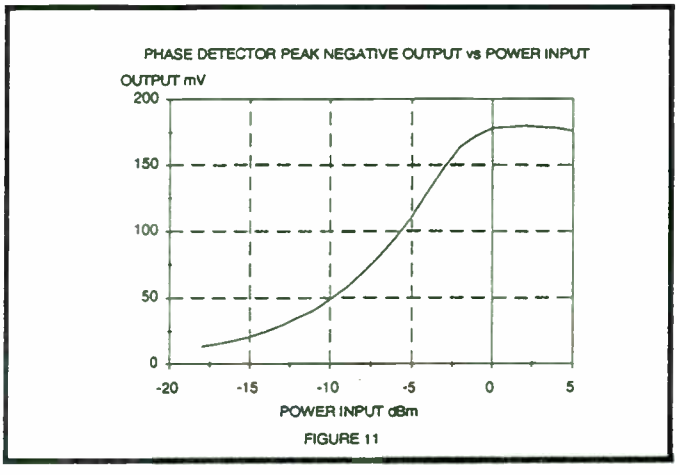


FIGURE 11

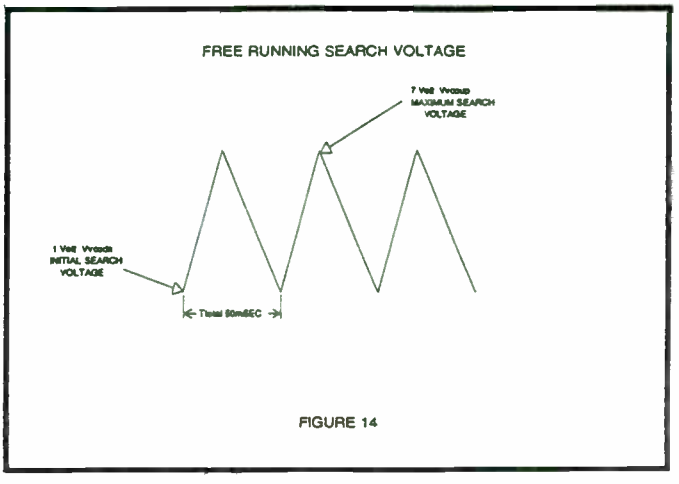


FIGURE 14

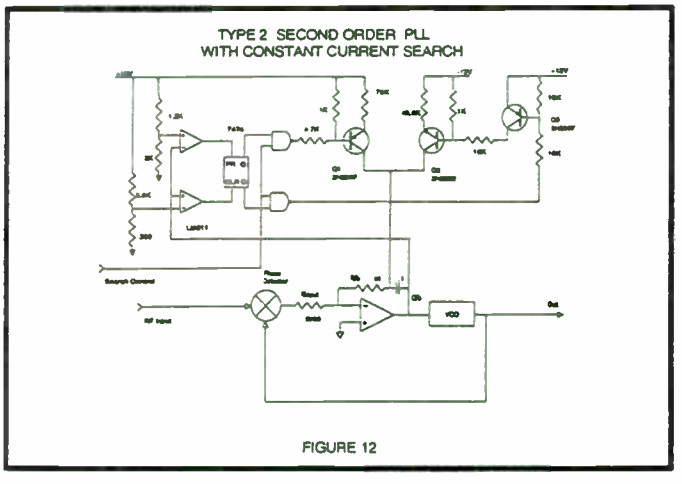


FIGURE 12

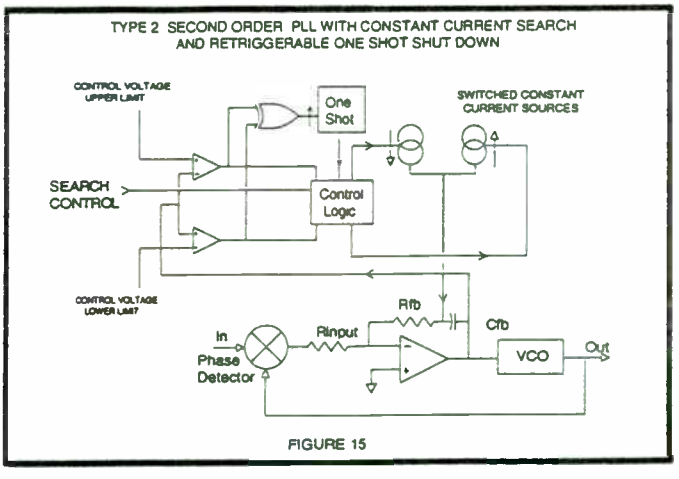


FIGURE 15

Analysis of Transversely Coupled SAW Resonator Filters Using Coupling of Modes Technique

by

V. Narayanan and S. M. Richie
University of Central Florida
Electrical and Computer Engineering Department
Orlando, Florida 32816-2450

ABSTRACT

This paper presents the Coupling of Modes (COM) analysis technique as applied to SAW device component elements, such as the SAW reflective grating and the interdigital transducer, which are then cascaded to analyze a class of SAW resonator filters called proximity coupled or waveguide coupled resonator filters. It also presents the technique to determine the number of transverse guided modes for a particular device structure employing a simplified waveguide model. The excitation and propagation of the modes can then be solved independently for each mode with its corresponding mode velocity using a PC based CAD tool written to implement COM solutions for the various SAW components. The proximity resonator is structured to have a coupling of resonant energy between two transducers placed in close longitudinal proximity to each other with a single reflective grating on each side. The number of transverse modes in a device structure is dependent on the acoustic beamwidth of the structure and for a device with small beamwidth there can be two dominant guided modes, the velocity difference of which results in a two pole response. In addition to the discussion of the analysis of proximity coupled SAW resonators, this paper presents the responses of other SAW structures obtained using a computer model "RESCAD" developed at the University of Central Florida.

INTRODUCTION TO SAW RESONATOR FILTERS

Surface acoustic wave devices are being utilized more frequently in high performance RF applications as these devices are continually offering higher performance with lower cost in smaller packages. A surface acoustic wave resonator is a high Q element which is used for frequency control and narrow band filter applications operating in a frequency range between 20MHz and 2GHz. The lower frequency limit is determined by the size of the device but at high frequencies fabrication tolerances and propagation losses are the limiting factors.

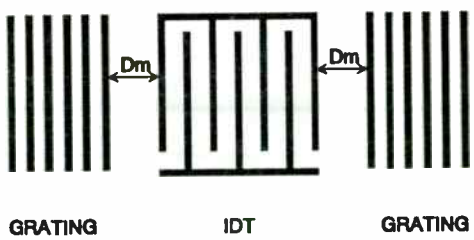


FIGURE 1 : One port resonator configuration

The commonly used resonators are the one port resonator, as shown in Figure 1, and the two port in line resonator, as shown in Figure 2. Other resonator filter structures include the acoustically coupled inline resonator filter and the proximity coupled resonator filter discussed in this paper.

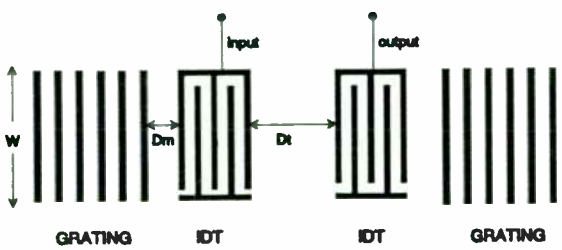


FIGURE 2 : SAW two port resonator

Quartz coupled resonator filters have provided a low cost solution for many VHF and UHF filter requirements. Over the past decade, SAW coupled resonator filters have seen a dramatic increase in their application as a key element in oscillator and RF filter circuitry because of their small size, low insertion loss and low relative cost. A typical proximity coupled resonator structure is shown in Figure 3.

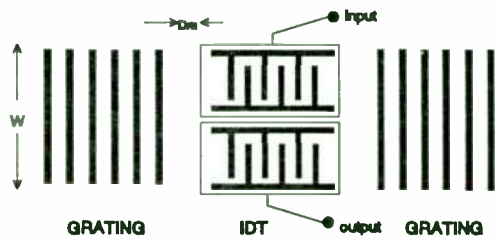


FIGURE 3: Transversely coupled SAW Resonator Structure

The proximity coupled or the guided mode resonator filter was first demonstrated in 1975 by H.F.Tiersten and R.C.Smythe [2], who described the coupling between resonators in terms of trapped energy modes of the surface wave guiding structure. The proximity of the transducers results in resonant energy in each of the two cavities being coupled together by virtue of the evanescent acoustic fields. Since that time a significant amount of SAW coupled resonator designs have been studied [1].

MODELING APPROACHES

Since the invention of SAW devices, a large variety of filters and resonators have been developed and used in signal processing applications. Many different device models have been derived to analyze and synthesize them. Physical effects like the excitation of different wave types, perturbations and scattering caused by different metallisation structures, wave diffraction, and other phenomena of wave propagation in the different materials have also been studied. Presently, there exists several device models which have become very popular in the analysis of SAW devices. These include the equivalent circuit model, the impulse response model, the coupling of modes model, and the waveguide model. The inline and cross field **equivalent circuit model** allows for the analysis of SAW devices including second order effects like that of internal reflections. These internal reflections result from the presence of the electrodes in the SAW transducer. As Rayleigh waves propagate under a SAW transducer, portions of the waves reflect from the electrodes. Analysis of SAW devices using this model requires cascading sections of the transducers.The curved electrostatic fields in between the metal fingers near the surface are replaced by a homogeneous electrical field distribution parallel or perpendicular to the surface [3]. Each periodic section of the device is represented by an equivalent circuit and analyzing a typical device would require cascading several hundred such equivalent circuits. As a result this method of modeling can be very numerically intensive. Moreover, due to its numerical nature the equivalent circuit model is not applicable to synthesis. The **impulse model** is an analytic model used to analyze SAW devices which lack any internal reflections. Analysis of SAW devices using the impulse model may be done in real time. The strength of the impulse model has been in its usefulness as a synthesis tool since closed form expressions for SAW device responses are possible. The generalized impulse response is closely related to Fourier transformation of the transducer weighting function and therefore this model is called the impulse model [4]. It is also based on the Green's function analysis of wave coupling and the principle of charge superposition [5]. The **coupling of modes** model is considered to be a hybrid of the equivalent circuit model and the impulse model and is an analytic model which includes the effects of distributed internal reflections. This model may be thought of as a generalized impulse model which includes internal reflections in its analysis or as an analytic equivalent of the equivalent circuit model. Since closed form expressions for the responses of SAW devices may be obtained using the coupling of modes model, it is applicable to synthesis as well. With the growth of SAW technology, a great deal of attention has been given to the application of the coupling of modes (COM) theory, better known for its contribution in the analysis of microwave and optical structures. It was successfully applied to the analysis and synthesis of SAW devices with small distributed reflections by P. V. Wright and H. A. Haus in 1980 [6]. Since then it has been used for various SAW device applications. It has demonstrated its usefulness as a simpler, less cumbersome design and analysis tool. The use of the COM theory provides efficient computations and good accuracy for the analysis of SAW devices, as compared to most field theory approaches which involves greater model complexity and can be more numerically inefficient. Lastly, the **waveguide model** was developed for microwave and optical waveguides [7] and has long been used for the calculation of wave propagation in SAW devices. The effect of waveguiding and transverse modes in SAW resonators was discussed by O.Schwelb, E.L.Adler and G.W.Farnell in 1977. A well known principle in wave propagation is that if a region with low wave velocity is surrounded by faster regions, the wave

is confined to the slow region. The best way to achieve this is to provide a thin metallic overlay in the wave propagation path which slows the wave and provides the desired wave guiding, as shown in Figure 4.

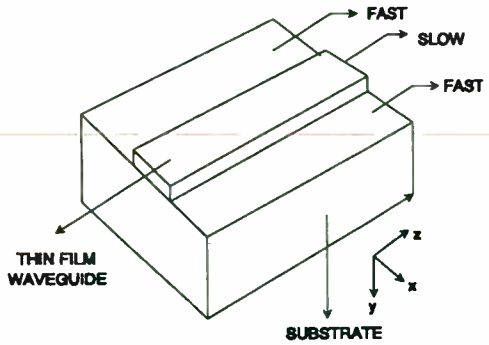


FIGURE 4 : SAW waveguide

Modes propagate with certain relative amplitudes in a waveguide. The model complexity required to account for waveguiding in the COM formalism or any other transducer model overshadows the usefulness of the model results and would be numerically cumbersome. The more prevalent approach is to account for the waveguiding independently from the SAW response model. It is best to use the waveguide model to determine the number of modes propagating in a particular device beamwidth for the analysis of a transversely coupled resonator filter. The response of each guided mode is then determined by the COM model.

ANALYSIS APPROACH

The technique used to analyze a SAW coupled resonator in this paper is based on the coupling of modes and waveguide model approach. Applying the coupling of modes formalism to a SAW structure, it is possible to calculate the required fundamental parameters from basic principles and conclude with the closed form solutions for the device response. The coupling of modes theory describes the acoustic field distribution in the resonator in terms of a forward traveling and a backward traveling wave. The coupling of modes formalism requires that the system be described by a set of N first order coupled differential equations governing the propagation of the N modes in the system. Each equation governs the propagation of a specific mode and describes how the remaining modes couple to it. In a transducer, as a Rayleigh wave propagates under the edge of an electrode, a portion of the wave reflects from the edge of the electrode. The portion of the wave that is reflected from the edge becomes part of the wave traveling in the opposite direction. Therefore, the reflections cause coupling of waves. Each of the waves propagating under the transducer will require that first order differential equations be derived describing the wave characteristics. Using this method, each of the SAW components can be analyzed independently. In a SAW device component there are four different types of transverse guided modes. These are the symmetric and antisymmetric trigonometric and the symmetric and antisymmetric hyperbolic modes. If the width of the transducer is under a certain limit only the trigonometric modes are guided by the device structure. Consider a uniform overlap SAW transducer, which can be Fourier decomposed to a sinx/x response in the frequency domain. Symmetric and antisymmetric modes are launched in the waveguide. The waves extend partially outside the guiding region and eventually die off as shown in Figure 5.

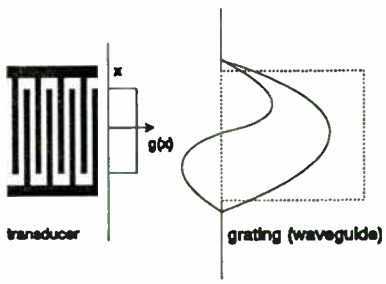


FIGURE 5

However, if another waveguide is placed next to the guiding region, the modes no longer die off and evanescent coupling results in the coupling of energy into the extended new waveguide and other modes are launched. In effect, for the analysis of proximity resonators, the two guiding regions are replaced by a single waveguide where the modes are supported.

COUPLING OF MODES ANALYSIS OF A SAW REFLECTIVE GRATING

The coupling of modes formalism describes the motion of the surface acoustic wave propagating across the substrate. It is derived from equations which are actually a set of first order wave equations. One of the main motivations for using the COM approach is its simplicity of application to SAW devices. Using the COM approach a closed form analytic description of periodic structures may be obtained which is of extreme importance especially when considering the synthesis of SAW devices. The reflective grating consists of 'n'

periodic strips with period 'p', strip width 'a_g' and height 'H_m' as shown in Figure 6.

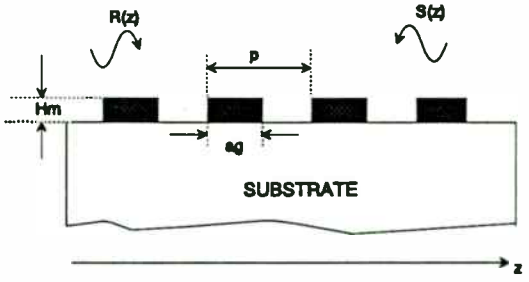


FIGURE 6 : A typical grating

A surface wave, U(z), can be broken into two propagating waves, the forward traveling wave a(z) and the reverse traveling wave b(z). The waves have slowly varying amplitudes, R(z) and S(z), respectively. The forward traveling wave, a(z), is assumed to be traveling in the +z direction and the reverse surface acoustic wave is assumed to be traveling in the -z direction. These waves are described by

$$a(z) = R(z)\exp(-jk_f z)\exp(j\omega t) \quad (1)$$

$$b(z) = S(z)\exp(+jk_f z)\exp(j\omega t) \quad (2)$$

where 'k_f' is the wave vector of the surface wave on a surface without electrodes, and 'w' is the angular frequency. Let 'k_g' be the wave number under the grating, so that

$$k_f = \frac{k_g}{2} \quad (3)$$

where,

$$k_g = \frac{2\pi}{p} \quad (4)$$

Assuming steady state time dependency, the factor exp(jwt) can be dropped in the analysis. An accurate description of the field distributions, e.g. displacements, would require an infinite number of "space harmonic" components, but in the COM approach only the two lowest order dominant harmonics are taken into account. The field may be written in the form of complex wave amplitudes and the local complex amplitude, U(z), can be approximated by superposition of the reverse and forward waves as stated below.

$$U(z) = R(z)\exp(-j\frac{k_g}{2}z) + S(z)\exp(j\frac{k_g}{2}z) \quad (5)$$

In the absence of the grating, the following differential equations are satisfied by the waves a(z) and b(z),

$$\frac{da(z)}{dz} + jk_f a(z) = 0 \quad (6)$$

$$\frac{db(z)}{dz} - jk_f b(z) = 0 \quad (7)$$

If a grating is placed onto the substrate, the periodic discontinuities disturb the propagation of the wave. The wave amplitude changes along 'z' because of the wave reflections from the grating strips and the phase velocity also changes since the free space velocity is higher than the velocity under metallized regions. The reflection mechanism therefore gives rise to coupling between the forward and the reverse waves. The strength of the coupling is determined by the form of the irregularities disturbing the propagation. Generally, a "mode" of propagation, means an infinite number of space harmonics characterizing the wave propagation. These space harmonics have different wavelengths at the same frequency and hence different phase velocities. In the COM approach, only one harmonic component is taken into account for each propagation direction. Therefore, the forward and backward waves are referred to as "modes". The presence of the grating changes the propagation constant by Δk so that

$$\frac{da}{dz} + j(k_f + \Delta k)a = jk_{21}b \quad (8)$$

$$\frac{db}{dz} - j(k_f + \Delta k)b = jk_{12}a \quad (9)$$

The periodic perturbations leads to power exchange between modes. R(z) couples to S(z) and S(z) couples to R(z). Using the differential form of the equations and assuming the change of amplitude per period of the grating is small, the law of conservation of energy requires

$$\frac{d}{dz}\{|R|^2 - |S|^2\} = 0 \quad (10)$$

since $R^2(z)$ and $S^2(z)$ correspond to the power carried by the "modes" R(z) and S(z) respectively. This condition leads to

$$k_{12} = -k_{21}^* \quad (11)$$

where the coefficient k_{12} is the coupling coefficient and is a periodic function of z. Generally speaking a complex Fourier series of the periodic function k_{12} of period 'p' is given by

$$k_{12} = \sum_{n=-\infty}^{\infty} K(n)\exp(jnk_g z) \qquad (12)$$

Substituting, (1) and (2) into (8) and (9) gives the following equations (13) and (14)

$$\frac{dR}{dz} = -j(k_f - \frac{k_g}{2} + \Delta k)R + j\sum_{n=-\infty}^{\infty} K(n)S\exp[j(n+1)k_g z]$$

$$\frac{dS}{dz} = j(k_f - \frac{k_g}{2} + \Delta k)S - j\sum_{n=-\infty}^{\infty} K^*(n)R\exp[-j(n+1)k_g z]$$

Of all n Fourier components, only the n=-1 term produces coupling between the forward and backward traveling waves that is independent of z. To simplify the equation, a "wave mismatch" parameter is used, called the detuning parameter as

$$\delta = k_f - \frac{k_g}{2} + \Delta k \qquad (15)$$

Using the detuning parameter and writing K(-1) as just K, the COM equations can be simplified as

$$\frac{dR}{dz} = -j\delta R + jKS \qquad (16)$$

$$\frac{dS}{dz} = j\delta S - jK^* R \qquad (17)$$

Now, solving these equations is straightforward and given by Elachi [8] as,

$$R(z) = c_1\exp[j(\delta - D)z] + c_2\exp[j(\delta + D)z] \quad (18)$$

$$S(z) = c_1\exp[-j(\delta + D)z] + c_2\exp[-j(\delta + D)z] \quad (19)$$

where an additional frequency variable D, the propagation constant which includes the detuning effects of the reflective grating, is included and is given as

$$D(\omega) = \sqrt{\Delta^2(\omega) - K_R(\omega)K_S(\omega)} \qquad (20)$$

$K_R(w)$ and $K_S(w)$ are the real and imaginary terms of the reflection (coupling) coefficient. The constant c_1 and c_2 are determined by boundary conditions. The total wave U(z)=a(z)+b(z) and its derivative with respect to position must be continuous. Figure 7 shows the behavior of the forward and reverse waves at the device boundaries. Since, the wave has to be a continuous function with respect to position, the boundary conditions are given to be

$$R(0) = 1 \qquad (21)$$

$$S(L) = 0 \qquad (22)$$

where L is the length of the grating and is given to be the product of the number of finger strips and the period of the grating structure, $L = np$.

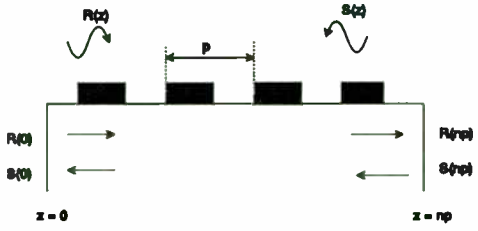


FIGURE 7 : Forward and Backward waves at the device boundaries

Solving for the coefficients c_1 and c_2 and substituting into (18) and (19) gives the amplitudes R(z) and S(z) to be

$$R(z) = \frac{D\cos[D(L-z)] + j\Delta\sin[D(L-z)]}{D\cos(DL) + j\Delta\sin(DL)} \qquad (23)$$

$$S(z) = \frac{jK^*\sin[D(L-z)]}{D\cos(DL) + j\Delta\sin(DL)} \qquad (24)$$

Figure 8, shows the relative amplitudes of the forward and backward traveling waves as a function of position z with fixed frequency.

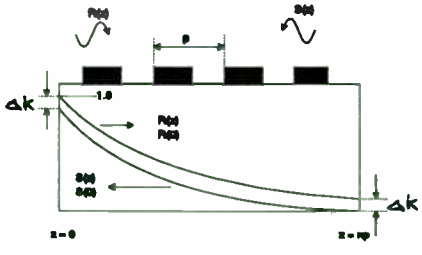


FIGURE 8 : Power coupling from incident to reflected wave

The magnitude of the incident wave R(z) decays exponentially along the grating due to the reflection of power into the backward traveling wave. The amplitude S(z) of the backward traveling wave builds up due to the power from the forward wave. This exponential decay behavior occurs only for a narrow band of frequencies, and only when the propagation constant D is real. Now,

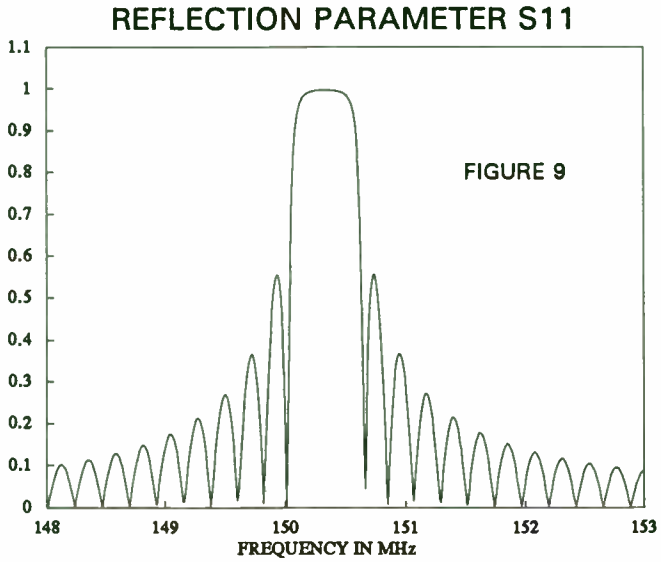
having obtained the wave equations, the scattering matrix characterization of the grating can be determined as

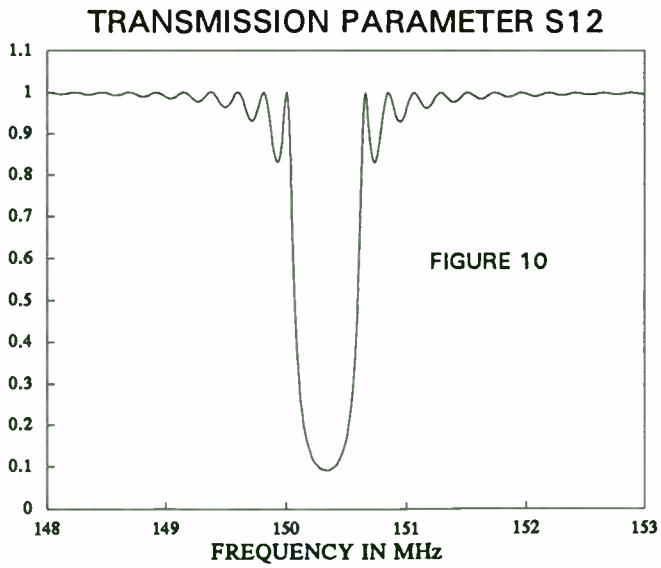
$$S = \begin{bmatrix} b(0) & a(np) \\ a(np) & b(0) \end{bmatrix} \qquad (25)$$

Therefore, making use of the solutions for the forward and backward waves as given in (23) and (24), the expressions for the scattering parameters is obtained to be

$$S_{11} = S_{22} = \frac{jK^* \sin(DL)}{D\cos(DL) + j\Delta \sin(DL)} \qquad (26)$$

$$S_{12} = S_{21} = \frac{(-1)^n D}{D\cos(DL) + j\Delta \sin(DL)} \qquad (27)$$





The grating can be analyzed separately and the scattering parameters can be obtained directly using the RESCAD computer model. The solutions for a typical grating are plotted in Figures 9 and 10.

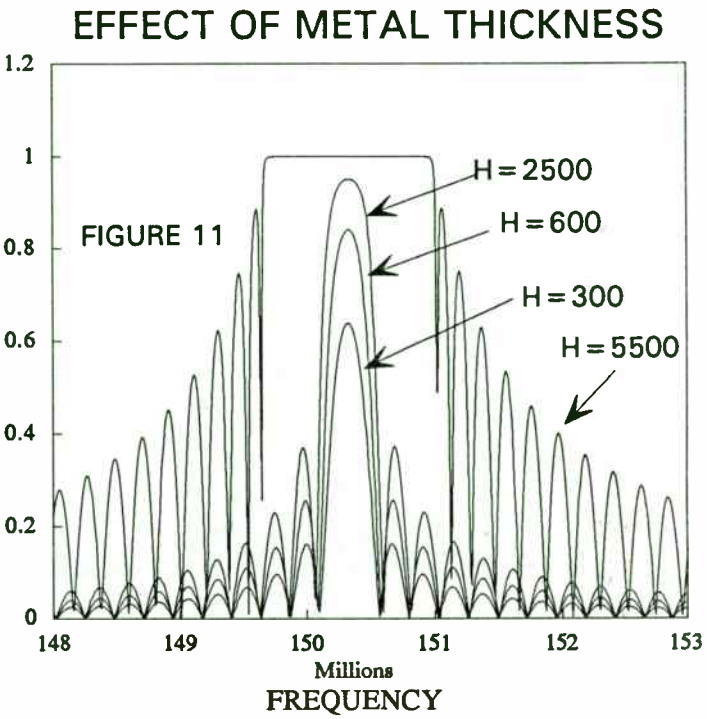


Figure 11 shows the effect of metal thickness on the grating reflection coefficient. The reflection coefficient is described by (26) where 'K' the reflection (coupling) coefficient is given by [9]

$$K = \left\{ \left[\frac{2\pi f}{v_f} \right] * \left[\frac{e_4 k^2}{2} + \frac{e_5 H_m}{L} \right] \right\} \qquad (28)$$

where e_4 and e_5 are material constants for the substrate, 'k^2' is the acoustic coupling coefficient of the substrate material and the other notations have the usual meanings. To show the variation on the reflection coefficient with respect to metal thickness, the number of grating electrodes was held constant in Figure 11. As can be observed from Figure 11, as the metal thickness increases, the stopband for the reflector is broader and the sidelobes are higher. A broader stopband increases the number of longitudinal resonance modes. The unwanted longitudinal modes occur near the edges of the stopband, and to reduce or eliminate these modes the transmission response of the reflective grating should be centered in the reflector stopband. As the metal thickness of the reflector increases, the reflectivity also increases. Figure 12, shows the effect of the number of

grating strips on the reflection coefficient which is done by maintaining a constant metal thickness.

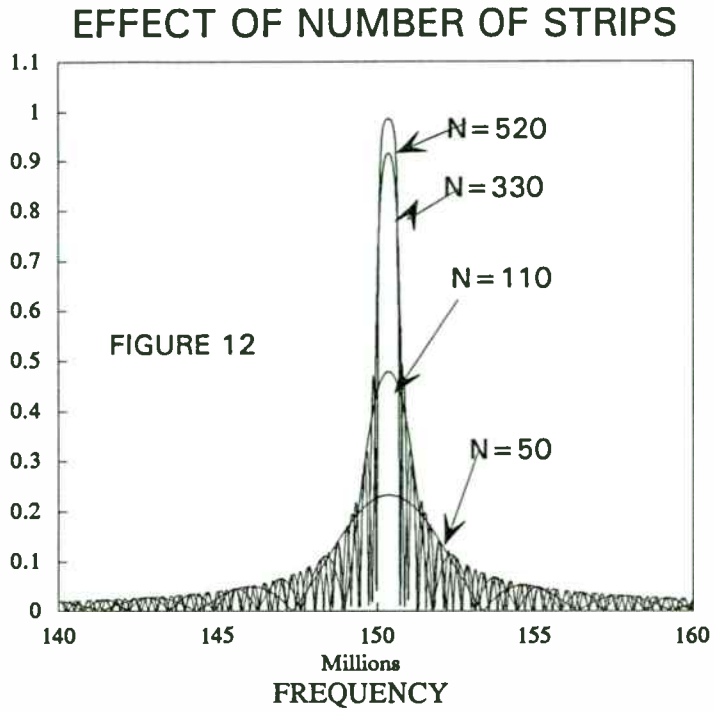


Figure 12 shows that the reflectivity increases with the number of electrodes. The passband gets narrower and the sidelobes become higher with an increase in the number of strips.

COUPLING OF MODES EQUATIONS FOR AN INTERDIGITAL TRANSDUCER

The transducer generates forward and backward propagating surface waves with complex slowly varying amplitudes, R(z) and S(z), that are coupled together, similar to a grating. The physical arrangement of the

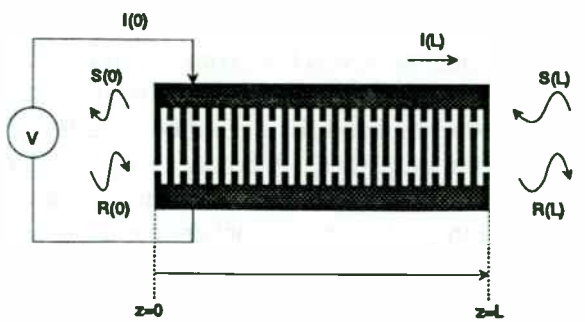


Figure 13 : IDT schematic

interdigital transducer with the forward and the reverse traveling wave is shown in Figure 13. The IDT is more complicated because of the combinations of elementary source transduction and internal reflections. A third port, an electrical one, is added. For a transducer, the basic COM equations of the grating will have to be modified by adding a new term called the transduction term. The transduction term describes the SAW excitation of a voltage V applied to a pair of electrodes. An equation is needed to represent the current drawn by the electrical port of the interdigital transducer. Within a uniform transducer, the effects of reflection and transduction can no longer be assumed to be independent. This relationship between the transduction and reflection weighting results directly from the structure of the transducer. The transduction period is twice the period of reflection. So, in order to get the COM equations for the IDT, the grating equations are altered by adding a transduction term and by using an additional equation to describe the relationship between the current and the voltage at the electrical port. Doing so yields,

$$\frac{dR}{dz} = jKS + j\xi V - j\delta R \tag{29}$$

$$\frac{dS}{dz} = j\delta S - jK^{*}R - j\xi^{*}V \tag{30}$$

The parameter ξ is the transduction coefficient. The electrical port can be described by

$$\frac{dI}{dz} = -2\xi^{*}jR + 2j\xi S + j\omega CV \tag{31}$$

where C is given as $C = C_S W$. C_s is the static capacitance per finger pair and W is the IDT aperture. R and S are amplitudes of the forward and backward waves. Equations (29), (30) and (31) are first order differential equations. The P-matrix notation has been used in the literature to present the results of a coupling of modes analysis. This is a very convenient approach since the solutions to the COM equations are dependent on the elements of the P matrix. In the P-matrix representation, the acoustic ports are treated as scattering ports and the electrical port as an admittance port. Figure 14 shows the three port scattering matrix description. The scattering and admittance properties of the three port junction can be written in matrix notation as

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{32}$$

The power incident on one port is distributed to the others depending on the coupling at the ports. A set of input waves is scattered into a set of output waves.

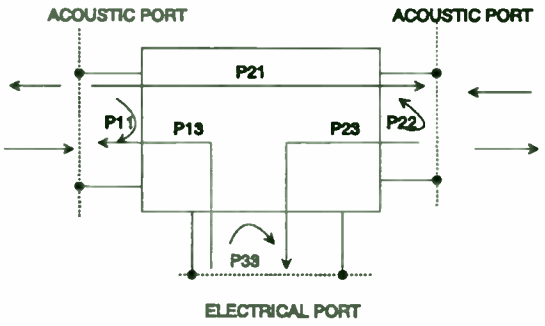


Figure 14 : P parameter description

The third port which is an electrical port is represented by the current drawn by the IDT and the terminal voltage. So equation (32) can be rewritten as

$$\begin{bmatrix} b_1 \\ b_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ V \end{bmatrix} \quad (33)$$

The IDT admittance is characterized by the P_{33} parameter. The P-matrix for a SAW transducer must satisfy three reciprocity relationships.[10]

$$P_{21} = P_{12};\ P_{31} = -2P_{13};\ P_{32} = -2P_{23}$$

Moreover, for a SAW transducer if loseless, the P-matrix will satisfy three unique energy conservation relationships. [10]

$$Re\{P_{33}\} = |P_{13}|^2 + |P_{23}|^2 \quad (34)$$

$$|P_{11}|^2 + |P_{21}|^2 = 1 \quad (35)$$

$$|P_{22}|^2 + |P_{12}|^2 = 1 \quad (36)$$

The advantages of using this representation is that it allows easier manipulation of the parallel electrical circuit and acoustically cascaded connection for multiple transducer/grating circuits. The elements P_{11}, P_{12}, P_{21} and P_{22} are identical to the scattering parameters developed for a grating. P_{33} is the admittance of the IDT as seen at the electrical port if no incoming waves are present and P_{13} and P_{23} are the voltage to SAW transfer elements. Applying B.P.Abbott's solutions [10] for the uniform IDT and expressing them in P-matrix form yields

$$P_{11} = S_{11} = \frac{jK^* \sin(DL)}{D\cos(DL) + j\Delta \sin(DL)} \quad (37)$$

$$P_{12} = S_{12} = \frac{(-1)^n D}{D\cos(DL) + j\Delta \sin(DL)} \quad (38)$$

$$P_{13} = jLB\left[\frac{\xi^* D\cos(\frac{DL}{2}) + j(K^*\xi + \Delta\xi^*)\sin(\frac{DL}{2})}{D\cos(DL) + j\Delta\sin(DL)}\right] \quad (39)$$

where $$B = \left[\frac{\sin(DL/2)}{DL/2}\right]$$

$$P_{22} = \frac{jK(-1)^{2n}\sin(DL)}{D\cos(DL) + j\Delta\sin(DL)} \quad (40)$$

$$P_{23} = jL(-1)^n B\left[\frac{\xi D\cos(\frac{DL}{2}) + j(K\xi^* + \Delta\xi)\sin(\frac{DL}{2})}{D\cos(DL) + j\Delta\sin(DL)}\right] \quad (41)$$

$$P_{33} = -j2\frac{X}{D^3}\left[DL - \frac{D\sin(DL) + j\Delta(1-\cos(DL))}{D\cos(DL) + j\Delta\sin(DL)}\right]$$

$$-2\frac{Y}{D^3}\left[\frac{1-\cos(DL)}{D\cos(DL) + j\Delta\sin(DL)}\right]$$

$$+j\left[\frac{3\omega C_F L/\Lambda_T}{3 + j\omega C_F R_F}\right] \quad (42)$$

where $X = K^*\xi^2 + K\xi^{*2} + 2\Delta|\xi|^2$ and

$$Y = \Delta(K\xi^{*2} + K^*\xi^2) + 2|K|^2|\xi|^2$$

Using the reciprocity relationship for the mixed P-matrix the remaining elements can be found. Using energy conservation the acoustic conductance can be expressed in terms of the waves leaving the SAW transducer.

$$G(\omega) = Re\{P_{33}\} = |P_{13}|^2 + |P_{23}|^2 \quad (43)$$

The interdigital transducer can be analyzed separately and the P-matrix parameters can be obtained using

"RESCAD". The typical P parameter responses are shown in Figure 15.

COM PARAMETERS

In order to use the coupling of modes technique, the various parameters used in the COM equations will have to be determined . The **detuning parameter** is also the "wave mismatch" parameter and is defined in (15) where 'k_f' is the wave vector of the surface wave on a surface without electrodes and is also the undisturbed propagation constant at the Bragg frequency, so that

$$\delta = \frac{2\pi f}{v_f} - \frac{\pi}{p} + \Delta k \qquad (44)$$

Δk is relatively small and can be neglected from the above equation. The **reflection coefficient** 'K' models the effects of the electrode strips on the surface wave, electrical loading, as well as the various mechanical loadings. A strip which is conducting, when placed on a piezoelectric substrate shorts out the tangential electric field associated with the wave. The electrical loading reduces the surface wave velocity and provides a mismatch of the gap regions and the electrode which have different wave impedances. The mechanical perturbations arise from different mechanical natures. Firstly, loading occurs due to the mass of the electrode metal which results from the difference in the mass densities of the metal strip and the piezoelectric substrate [9]. Secondly, there is loading due to elasticity of the strips. To reduce these mechanical disturbances, a metal film that has elastic properties and mass density similar to the substrate is desired for the grating strips. The reflection coefficient is given by equation (28) and is dependent on material constants. There are two coupling of modes parameters used to model **transduction**. These are the transduction strength and the transduction phase. The transduction process is common to all the SAW device analysis methods. There is a fundamental difference between the impulse model and the coupling of modes model in that the impulse model assumes that the reflectivity from the transducer electrodes is zero. As a result K must be zero. This effectively uncouples the two acoustic modes. Comparing the impulse model solutions for P_{13} with the COM model solution yields

$$\xi = |k| \in_s (\infty) \frac{V_F(k)}{L_t} \sqrt{\frac{\omega W \Gamma_s}{2}} \exp(j\Phi_T) \qquad (45)$$

where 'k' is the wave number, $\in_s$ is the electrostatic permittivity of the substrate, Γ_s is the SAW coupling coefficient defined by Morgan [11] , $V_F(k)$ is the Fourier transform of the elemental potential of the transducer structure and L_t is the transduction period. An expression for the **static capacitance** per finger pair, was given by Engan[12] and is computed by calculating the total charge per unit length of each electrode.

$$C_S = (\in_0 + \in_p) \frac{K(\xi)}{K(\sqrt{1-\xi^2})} \qquad (46)$$

where $\in_0$ is the dielectric permittivity and $\in_p$ is the zero stress permittivity.

TRANSVERSE MODE EQUATIONS

In order to determine the number of modes in a device structure, the dispersion relationship dependent on the transverse boundary conditions must be solved. In other words, the number of modes has to be solved by taking into account the different velocities in the different metallized regions such as the bus bars, finger region and free surface. Each mode can be viewed as a plane wave propagating at an angle to the direction along the guide, which is reflected at the waveguide boundaries, following a zig zag path [13].

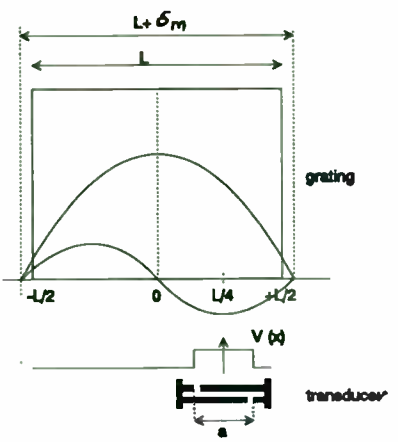


Figure 16 : Occurrence of modes

The basic equations for the mode analysis of a device structure propagating symmetric and antisymmetric modes as shown in Figure 16 is considered. The equation for the rectangular function is given as

$$V(x) = A\,rect\left(\frac{x - L/4}{a}\right) \qquad (47)$$

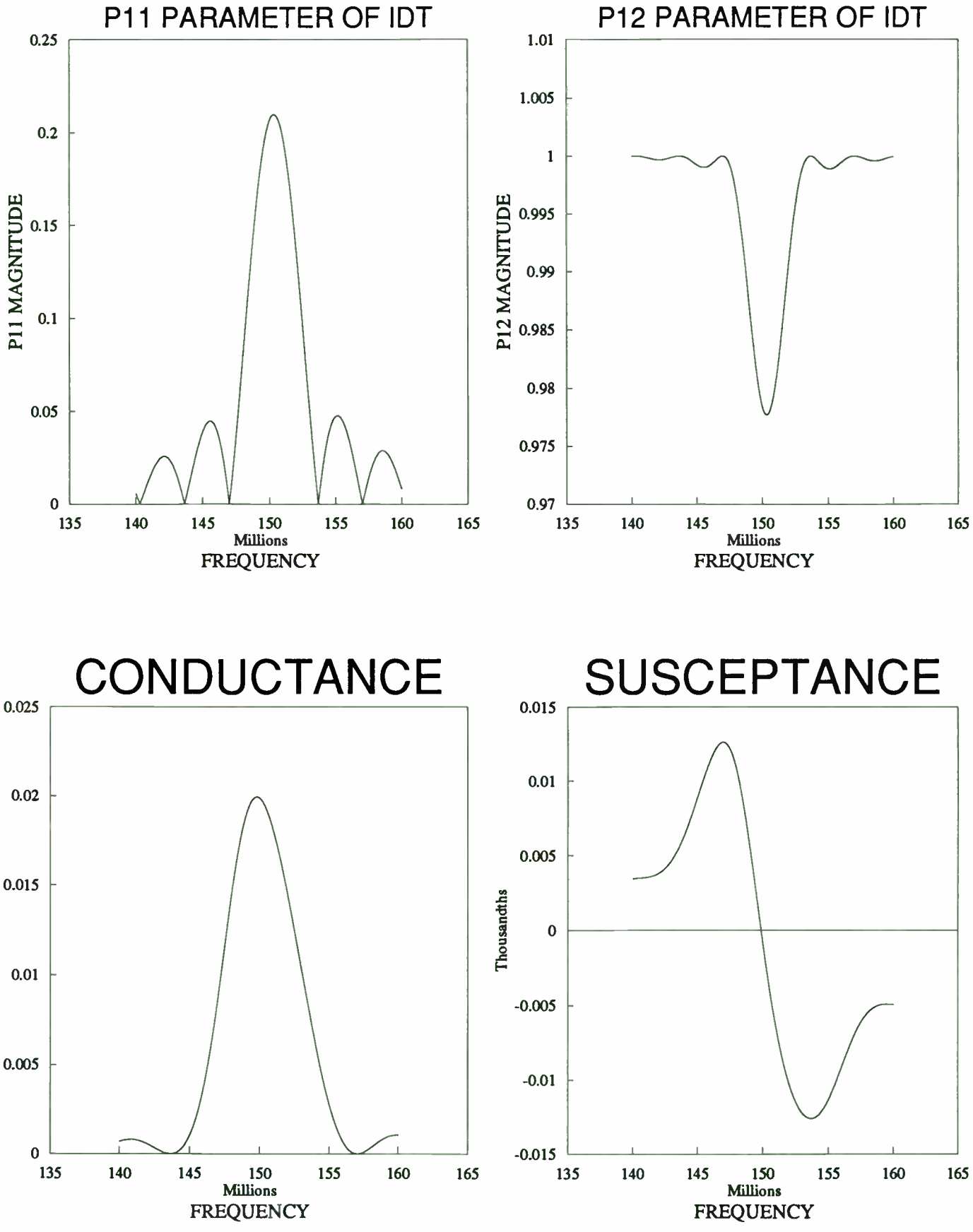


FIGURE 15: COM solutions for an IDT

This rect function V(x) is the sum of all mode profiles $f_m(x)$, where $f_m(x)$ is defined separately for symmetric modes as

$$f_m(x) = a_m \cos \frac{(2m-1)\pi x}{L+\delta_m} \tag{48}$$

and for the antisymmetric modes as

$$f_m(x) = b_m \sin \frac{2\pi m x}{L+\delta_m} \tag{49}$$

where, 'L' is the width of the transducer region, 'm' is the mode number , a_m and b_m are the normalized mode amplitudes and δ_m is the effective widening of the mode shapes as a result of the Goos Haechen effect [7].Since, V(x) is the sum of all mode profiles,

$$V(x) = \sum_{m=1}^{\infty} a_m \cos \frac{(2m-1)\pi x}{L+\delta_m} + b_m \sin \frac{2\pi m x}{L+\delta_m} \tag{50}$$

The wave number of the guided wave k, is related to that of the unguided wave number by

$$k^2 = \beta_m^2 + \left(\frac{m\pi}{L+\delta_m}\right)^2 \tag{51}$$

In order to use (51), the effective widening parameter δ_m must be known, which can be determined by

$$\delta_m = \frac{L 2\theta_m}{m\pi - 2\theta_m} \tag{52}$$

where
$$\sin(\theta_m) = \frac{m\pi - 2\theta_m}{k(L+\delta_m)\sqrt{1 - v_m^2/v_f^2}} \tag{53}$$

The relation between the wavenumber of the guided wave and the unguided wavenumber becomes

$$\frac{\beta_m^2}{k^2} = 1 - \left[(1 - \frac{v_m^2}{v_f^2})\sin^2\theta_m\right] \tag{54}$$

where v_f and v_m are the velocities in the substrate and the metal respectively and β_m is the modal wavenumber. For a transducer electrode region the wave is slowed down by $\Delta v/v$ because of the perturbations of the wave by the electrodes. It is assumed that this velocity ratio does not change for waves at an incident angle to the electrodes, since the effective metallisation ratio of the electrodes to free space is constant for all waves propagating at different angles to the electrodes. Equation (50) is used to calculate the normalized mode amplitudes, where the transverse excitation function V(x) is expressed for an active fingerpair as the decomposition into a modified non-orthogonal Fourier series with Fourier coefficients a_m and b_m given by

$$a_m = \frac{1}{L+\delta_m}\left\{\int_{-L/2}^{L/2} |V(x)| \cos\left[\frac{(2m-1)\pi x}{L+\delta_m}\right]\right\} dx \tag{55}$$

$$b_m = \frac{1}{L+\delta_m}\left\{\int_{-L/2}^{L/2} |V(x)| \sin\left[\frac{2m\pi x}{L+\delta_m}\right]\right\} dx \tag{56}$$

$|V(x)|$ is the amplitude of the rectangular function. The decomposition into modes uses alternate sine and cosine mode shapes which satisfy the boundary conditions. Therefore the antisymmetric modes will have to be close to zero near the edges of the slow region. In order to calculate the number of modes, the simultaneous equations (52) and (53) have to be solved for δ_m and θ_m as suggested in [14]. Once these parameters are determined, the wave numbers can be determined using (54) and the amplitude profiles can be determined from equations (48) and (49). In order to calculate the number of modes in a device structure, the grating aperture and the ratio of the velocities $\Delta v/v$ should be known. Once the number of modes that the device structure can support is known,, the corresponding mode velocities can be calculated to analyze the device structure. Each mode propagates with a different velocity and have different COM parameters. The mode velocity for each mode can be determined by solving for the wavenumber associated with that mode using (54) given the substrate velocity is known.

CASCADING COMPONENTS

In order to analyze SAW resonator filter structures, the basic SAW components will have to be cascaded efficiently by modeling them as 3X3 P-matrices, with the appropriate terms forced to zero for the reflective grating and the delay. The delay represents the distance between the reflective grating and the interdigital transducer. For the simple configuration of cascading two structures, each structure can be considered as a separate three port device having two acoustic ports and one electrical port. When cascaded, the acoustic ports are effectively in series and the electrical ports are in parallel as shown in Figure 17. The P-matrix of the cascaded pair becomes [4]

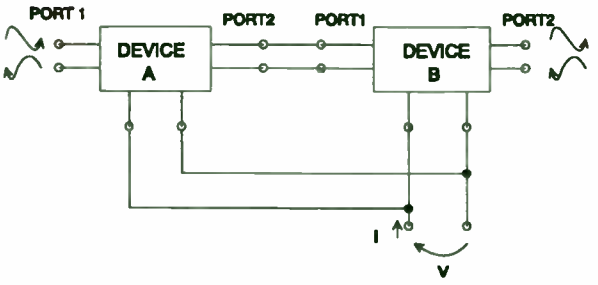


FIGURE 17 : Cascading P matrices

$$P_{11} = P_{11A} + P_{11B}\left[\frac{P_{21A}P_{12A}}{1-P_{11B}P_{22A}}\right] \quad (57)$$

$$P_{12} = \frac{P_{12A}P_{12B}}{1-P_{11B}P_{22A}} \quad (58)$$

$$P_{13} = P_{13A} + P_{12A}\left[\frac{P_{13B}+P_{11B}P_{23A}}{1-P_{11B}P_{22A}}\right] \quad (59)$$

$$P_{22} = P_{22B} + P_{22A}\left[\frac{P_{12B}P_{21B}}{1-P_{11B}P_{22A}}\right] \quad (60)$$

$$P_{23} = P_{23B} + P_{21B}\left[\frac{P_{23A}+P_{22A}P_{13B}}{1-P_{11B}P_{22A}}\right] \quad (61)$$

$$P_{33} = P_{33A} + P_{33B} + P_{32A}\left[\frac{P_{13B}+P_{11B}P_{23A}}{1-P_{11B}P_{22A}}\right] + P_{31B}\left[\frac{P_{23A}+P_{22A}P_{13B}}{1-P_{22A}P_{11B}}\right] \quad (62)$$

The above technique is used to cascade first the left most grating to the delay of the resonator in Figure 1. The equivalent P matrix is then cascaded with the IDT P-matrix, and so on, until a single equivalent P matrix is obtained for the device. A proximity resonator is analyzed by cascading the SAW components the same way. For the two port resonators shown in Figure 2, the IDT's acoustic ports are in series through a delay representing the distance between the transducers, but their electrical ports are not connected at all. If the outer acoustical port of each IDT is terminated in its characteristic impedance, the device can be reduced to an equivalent two port, with the IDT electrical ports being the ports of interest. The effects of the grating with the associated delay is accounted for by cascading each IDT to its grating as in the case for a one port resonator. The two equivalent P-matrices are converted to an equivalent two port admittance matrix which are then converted to the corresponding S-parameters [15].

$$S_{11} = \frac{(1-Z_1^*Y_{11})(1+Z_2Y_{22})+Y_{12}Y_{21}Z_2^*Z_2}{(1+Z_1Y_{11})(1+Z_2Y_{22})-Y_{12}Y_{21}Z_1Z_2} \quad (63)$$

$$S_{12} = \frac{-2\sqrt{R_1R_2}\,Y_{12}}{(1+Z_1Y_{11})(1+Z_2Y_{22})-Y_{12}Y_{21}Z_1Z_2} \quad (64)$$

$$S_{21} = \frac{-2\sqrt{R_1R_2}\,Y_{21}}{(1+Z_1Y_{11})(1+Z_2Y_{22})-Y_{12}Y_{21}Z_1Z_2} \quad (65)$$

$$S_{22} = \frac{(1-Z_2^*Y_{22})(1+Z_1Y_{11})+Y_{12}Y_{21}Z_1Z_1^*}{(1+Z_1Y_{11})(1+Z_2Y_{22})-Y_{12}Y_{21}Z_1Z_2} \quad (66)$$

EQUIVALENT CIRCUIT REPRESENTATION

The proximity device structure can be represented by an equivalent circuit keeping in mind that the key element is the admittance parameter at the electrical port. Figure 18 shows the equivalent circuit for a device supporting two modes.

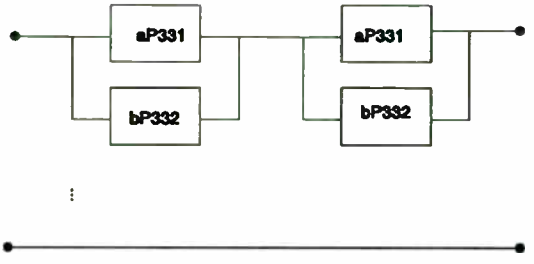


FIGURE 18 : Equivalent Circuit for a 2 mode device structure

The coefficients 'a' and 'b' represent the couplin between the modes and 'P331' and 'P332' represent the admittance's for the first mode and the second mode respectively. Figure 19 shows the equivalent circuit for a structure propagating N modes. The equivalent circuit

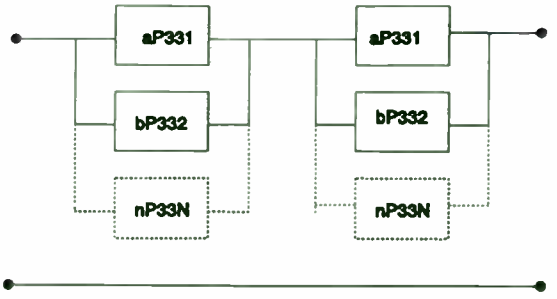


FIGURE 19 : Equivalent Circuit for a 2 mode device structure

for structures propagating various modes can be represented as a two port admittance parameter network as shown in Figure 20, from which the S-parameters for the proximity device structure of interest can be computed.

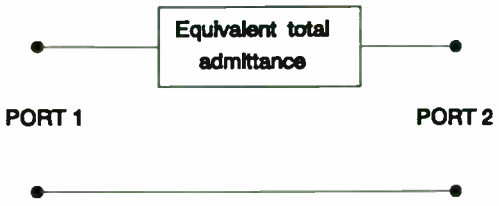


FIGURE 20 : Equivalent Circuit

The two port equivalent network can be shown as a circuit excited by a source and terminated at a normalizing impedance Z_o , as shown in Figure 21.

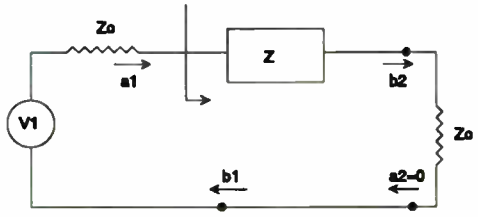


FIGURE 21 : two port network for s-parameter determination

From Figure 21, the S-parameters for the network are determined to be [16]

$$S_{11} = \frac{Z}{Z+2Z_o} \qquad (67)$$

$$S_{21} = \frac{2Z_o}{Z+2Z_o} \qquad (68)$$

Knowing the admittance's of the various modes and the coupling associated with them, the S-parameters for the device can be calculated.

COMPUTER MODEL

A computer program using the C programming language is written to implement the coupling of modes analysis technique. The program is capable of modeling the basic SAW components, the reflective grating and the interdigital transducer which are essential to model SAW resonator filters. The package "RESCAD" developed at the University of Central Florida is also capable of modeling some resonator structures like the one port resonators, the inline two port resonator structures and a two port SAW guided mode resonator filter. The "RESCAD" architecture is shown in Figure 22. The model sets an upper bound of 10000 data points

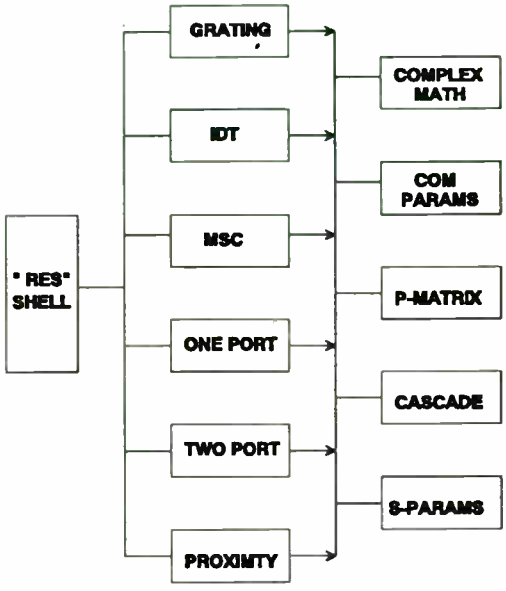


FIGURE 22 :"RESCAD" architecture

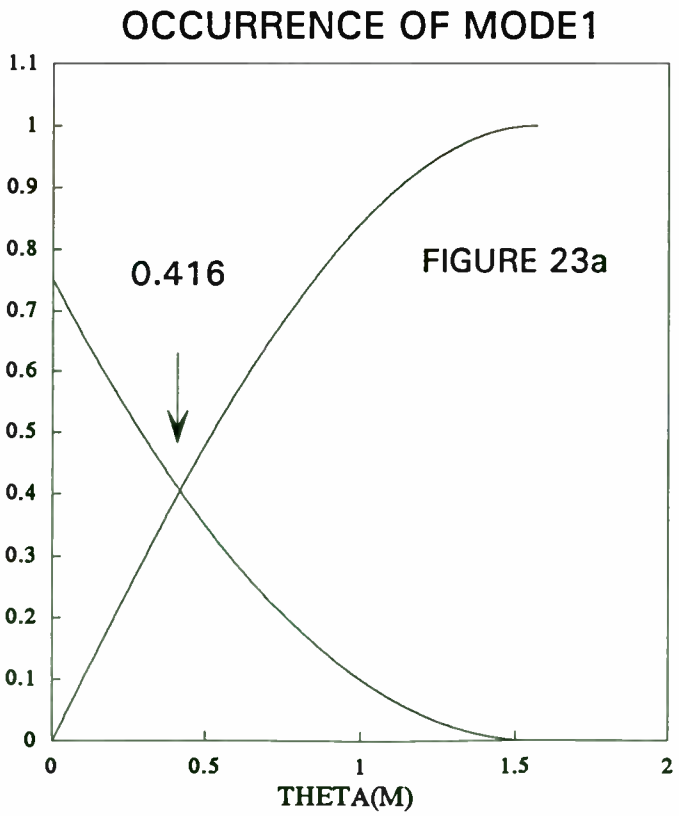
for the user specified frequency range. After choosing the structure of analysis, the user is prompted to open a file to store the output data, the geometry of the device and the analysis range. The output of RESCAD, the S-parameter data of the chosen structure. In addition, in order to model proximity resonators, the program takes into account the number of modes the device is capable of supporting, the corresponding mode velocities and the mode coupling coefficients. For each mode, the program cascades the P-matrices of the SAW components including the delay and finally calculates the S-parameters of the overall device structure from the admittance parameters of the various modes. The S-parameter output file can then be easily imported onto a commercially available spreadsheet for plotting.

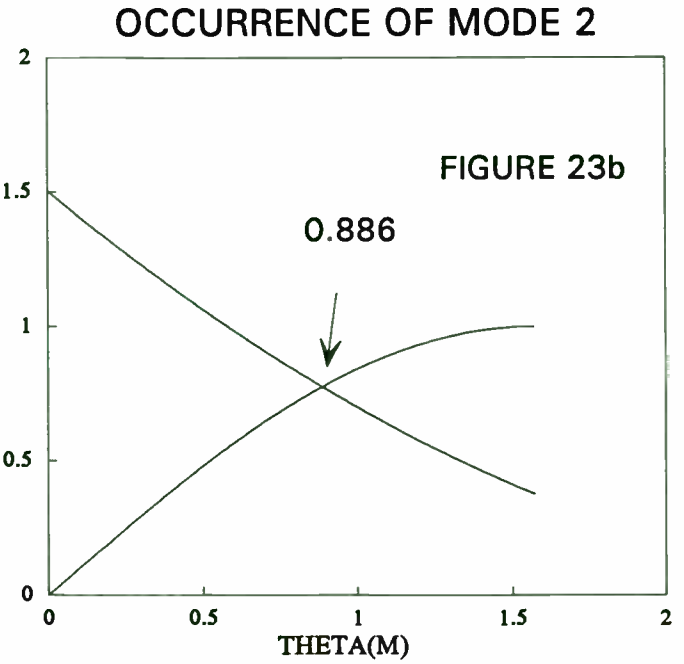
EXAMPLE

To check the validity of the computer model "RESCAD" written to implement the coupling of modes solutions, a sample example with certain assumed device specifications listed in the following table is executed.

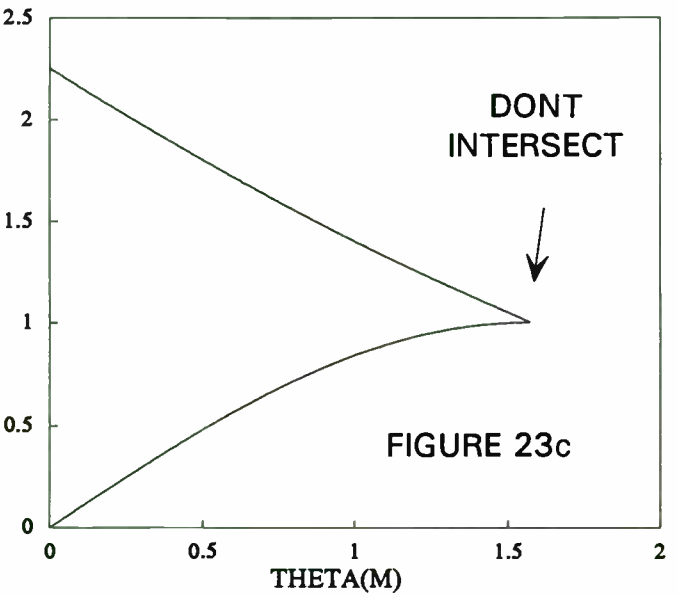
Number of grating strips	650
Number of IDT fingers	45
period of the grating	10.5um
period of the IDT	10.5um
Center Frequency	150MHz
Metal Thickness	1800A
Width of the Grating	405um
Resistivity	0.2
Substrate Assumed	Quartz
Velocity	3157m/s
Delay	13.125um
Width of the electrodes	5.25um

In order to calculate the number of transverse guided modes in the device structure equation (53) has to be solved. One way of solving this is by plotting both the sides of equation (53) from which the number of modes can be determined as shown in Figure 23. Figure 23 shows that there are just two transverse guided modes possible in this proximity device structure.

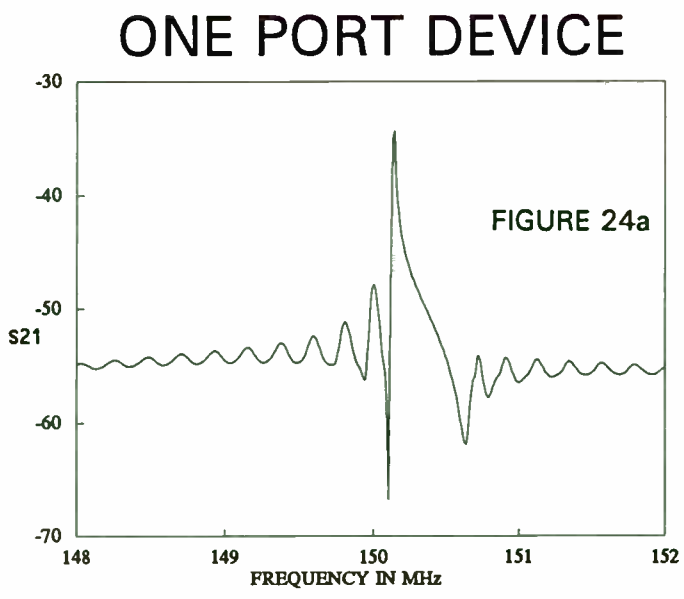


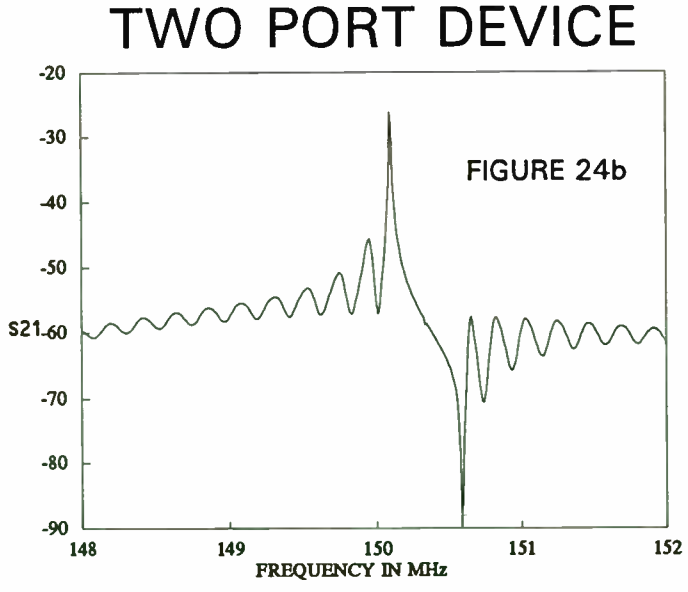


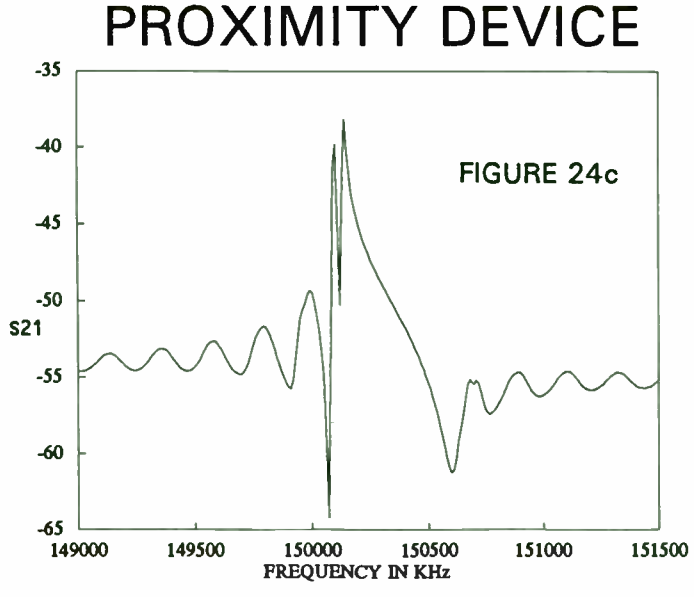
MODE 3



Once θ_m is known, the wavenumber of the guided modes can be determined from (54). The velocities of the two modes can then be determined knowing the substrate free surface velocity. In addition, the mode coupling coefficients can also be determined using (55) and (56). The mode velocities are found to be 3157.3m/s and 3158.13m/s for the first and second modes respectively. The responses for the various structures are given below in Figure 24.







REFERENCES

[1] P.V.Wright "A review of SAW resonator filter technology" 1992 IEEE Ultrason. Symp. Proc.

[2] H.F.Tiersten and R.C.Smythe ," Guided Acoustic surface wave filters", 1975 IEEE Ultrason. Symp. Proc.,pp 293-294.

[3]W.R.Smith et al.,"Analysis of interdigital transducers by use of an equivalent circuit model",1979 IEEE Trans.

[4] C.S.Hartmann and B.P.Abbott," Generalised impulse response model for SAW transducers including effects of electrode reflections", 1988 IEEE Ultrason. Symp. Proc., pp40 - 45.

[5] W.R.Mader, H.R.Stocker,"Extended impulse model for the design of precise SAW filters on quartz", 1982 IEEE Ultrason. Symp. Proc. pp 29-36.

[6] H.A.Haus and P.V.Wright, " Analysis of grating structures by Coupled mode theory",1980 IEEE Ultrason. Symp. Proc.

[7] N.S.Kappany and J.J.Burke, " Optical waveguides ", Academic Press 1972, pp 74-80.

[8] C.Elachi,"Waves in active and passive periodic structures:A review",Proceedings of the IEEE,Dec 1976.

[9] T.Thorvaldsson,"Analysis of SAW resonators".

[10] B.P.Abbott, " Coupling of modes model for SAW transducers with arbitrary reflectivity weighting", Ph.D. dissertation, University of Central Florida 1989.

[11]D.P.Morgan,"Surface acoustic wave devices for signal processing".

[12] Engan.H.,"Excitation of elastic surface waves by spatial harmonics of interdigital transducers",1969 IEEE Trans on Electron Devices. pp 1014-10

[13]S.Jungling," The impulse model including waveguiding effects using modal analysis" 1991 IEEE Ultrason. Symp. Proc.

[14] S.Datta ,"surface acoustic wave devices", 1986 Prentice Hall, Englewood Cliffs.

[15] Pozar David .M.,"Microwave Engineering" Reading, Mass.: Addison-Wesley, 1990

[16]G.Gonzalez ," Microwave transistor amplifiers, design and analysis", 1984 Prentice Hall, Englewood Cliffs

[17] Robert P.Burke, " A coupling of modes analysis of SAW resonators", University of Central Florida.

[18] M.Tanaka,T.Moriata et al, " Narrow bandpass filter using double mode SAW resonators on Quartz" , IEEE 38th Annual Frequency Control Symposium,1984.

Designing Microwave Circuits for Geosynchronous Space Applications

Ronald Ogan
TRAK Microwave Corporation
4726 Eisenhower Blvd.
Tampa, Florida 33634

INTRODUCTION

Today's world is a blur of information, some of which is transmitted via microwaves to and from earth stations through satellite electronics systems. Most communication satellites are placed in geostationary orbits, which have a radius of 35,880 km (22,300 miles) in the equatorial plane with a period equal to the rotation of the earth about the poles. A satellite in a geosynchronous orbit appears stationary over one point on the equator. After launch via rocket or Space Shuttle, the satellite is relocated from a low parking orbit into the final orbit by a maneuver called a Hohmann transfer. Thrusters provide 3-axis stabilization over the projected operating life of 10 years or longer (for INTELSAT series). Telemetry and Command/Control beacons provide attitude and data required to efficiently manage the satellite from earth stations. INTELSAT-6 , by using advanced digital and modulation techniques can carry 120,000 telephone circuits and 3 television channels. The satellite uses TDMA(Time Division Multiple Access) technique to transmit in the 6-4 GHz (C-Band) and 14-12 GHz (Ku-Band) frequencies.

TECHNICAL OBJECTIVES

Designing microwave circuits for the geosynchronous space application requires pushing the envelope of device and assembly knowledge to the limits. Since the reliability of the design is paramount, the major aspects of the design are carefully analyzed: device selection, materials selection, assembly processes and compliance testing. These areas will be discussed in the sections to follow.

SPACE ENVIRONMENT

Space is a harsh mistress which includes radiation hazards, materials outgassing, and the severe limitation that electronics must function flawlessly for 10 or more years without repair. Also, another prerequisite to operating in a stable orbit in space is the survival of launch vibrations. Temperature variations are typically controlled by the satellite rotation from -25 C to +65 C. Exposure of the electronic circuits to temperature variations and radiation depends upon the location on the spacecraft platform. Microwave beacons which are mounted on the external sections are subjected to the environmental extremes.

DEVICE SELECTION/ QUALIFICATION

Active and passive components are selected to the maximum extent possible from established reliability devices as listed in MIL-STD-975. However, few microwave components appear in the military standard so the next choice is to find devices with space heritage. Semiconductor vendors such as NEC, Motorola, and National Semiconductor have listings of space programs on which their devices have been flown. The device selection process proceeds interactively through the initial design process. The initial target device specifications are defined in the initial design and changes are made as required based upon a combination of the worst case circuit analysis and the breadboard testing of the circuits.

Devices are qualified for use on flight circuits by passing Environmental Stress Screening (depending upon the starting reliability of the devices: "R" .1% failure rate, or "S" .01% failure rate) and a Destructive Part Analysis (such as per MIL-STD-1580). Other inspection criteria may be imposed depending upon the specific program requirements such as Scanning Electron Microscope, SEM, examination of the device wafer, Particle Impact Noise Detection, PIND, of the packaged devices, and radiography of each packaged device. Additional lot sample tests may be required such as destructive testing of lead strength and lead solderability.

DESIGN CRITERIA

Designing microwave circuits to meet stringent End-of-Life Tolerances requires determination of the circuit variables, their effect on the circuit output parameters and the

predicted aging variations over the environmental conditions. For oscillator circuits, the worst case analysis can be divided into the following sections: Frequency Stability, Amplitude Stability, Spurious frequencies, radiation effects and EMI/EMC effects. The circuit configuration which will be discussed is shown in Figure 1 and is typical of microwave circuits used in satellites for command receivers and telemetry beacons.

For oscillator circuits, the components which can directly affect the frequency are usually limited to the first three stages of the chain. Amplifier and multiplier stages which follow can affect the amplitude but cannot affect the oscillator frequency. In designing long life (10 year life), microwave circuits, strict derating guidelines must be met such as limiting the transistor power dissipation to 20% of the manufacturer's rating. When the impedance match between stages is tuned during the initial alignment process, amplitude peaking occurs. Worst case circuit analysis must be performed to establish limits for the transistor bias currents, capacitor, and inductor values as required for achieving the specific RF output limits. Figure 2 shows a non-linear simulation using EESOF™ Libra™ for Windows for a X3 multiplier with +13 dBm input at 452 MHz. Note that the 3rd harmonic is approximately at 0 dBm and the 2nd harmonic is at -2 dBm, 904 MHz. Figure 3 shows the variation of the first and second harmonic for the output amplifier with different RF input levels. Bandpass filters are used to reject the unwanted harmonics and spurious signals which result from the multiplication process.

EOL changes must be calculated for each component parameter to determine the effect upon the circuits critical parameters such as frequency stability, RF power stability and minimal spurious at the output port. EOL drift characteristics are often stated as guidelines by the program. Device manufacturer's collect and report aging data that is taken during qualification tests to the National Aeronautics and Space Administration, NASA in addition to reporting to other government agencies. Component changes over life effectively detune the resonant and impedance matching circuits which degrades the output waveform. Aging effects on the bandpass filters which are used to reject spurious frequencies result in broader skirts with higher insertion losses. As the inductors and capacitors age as a result of time and exposure to the space environment, the transfer characteristics are shifted and flattened (Q of the filters is reduced). The total parameter variation, P_{EOL} is determined:

P_{EOL} = Po + delta Pm + delta R
where:
Po = the initial Beginning-of-Life tolerance
Delta Pm = the calculated parameter change based on mission life and operating temperature
Delta R = the effects of radiation upon the device parameter.

Figure 4 shows a typical transistor multiplier configuration. Using Libra ™, a sensitivity may be performed manually or by Monte Carlo simulation to determine how variations of each component within the tolerance range can effect the harmonic power levels. The multiplier circuits are biased class B to allow them to turn on with the application of the input RF signal. The Libra™ simulator performs reasonably well against measured data but caution must be exercised when the multiplier output power increases directly as a function of the DC bias condition.

Crystal oscillators have aging characteristics that are set by the stability of the crystal. Recent advances in SC-cut crystals have resulted in frequency variations < .5 ppm per year for 100 MHz, fifth overtone crystals. The factors affecting frequency stability are as shown:

Initial set accuracy (may be adjusted out)	± 1.0 ppm
Aging over 10 years in space	± 5.0
Temperature variations	± .5 ppm
Power Supply variation	± .1 ppm
Shock/vibration	± .1 ppm
Radiation exposure	± .1 ppm
TOTAL	± 6.8 ppm

The initial set accuracy may be eliminated during the alignment tuning process by pulling the crystal on frequency using the circuit resonant components. However, SC-cut crystals, because of the their excellent long term aging characteristics, have a limited pulling range. Preconditioning of the crystal and monitoring of the aging trends is very important to insure long term stability. Initially, the crystal frequency rate of change is rapid until a stable plateau is reached.

CONCLUSION

Microwave circuits have been used in low altitude space programs such as the weather satellites for over 30 years beginning with TIROS-1 in 1960. Circuit complexity has increased steadily and now includes many requirements for telecommunication satellites in geosynchronous orbits.

Time Division Multiple Access technology has driven the need for more accurate clock oscillators to control the frequency channels. The future for geosynchronous applications such as Global Positioning Systems is a bright rising star in today's relatively flat technological landscape.

Non-linear circuit analysis tools such as EESOF™, Libra™ provide insight into how the circuits can be analyzed and improved to meet the stability demands. Device, materials, and assembly processes are improving reliability for missions that are now approaching 20 years in length. Long life programs require methodical approaches to device and material selection with the emphasis on reliability and design margins. Device environmental screening and strict operational derating further enhance the mission performance.

[Figure 1 appears on the following page. Ed.]

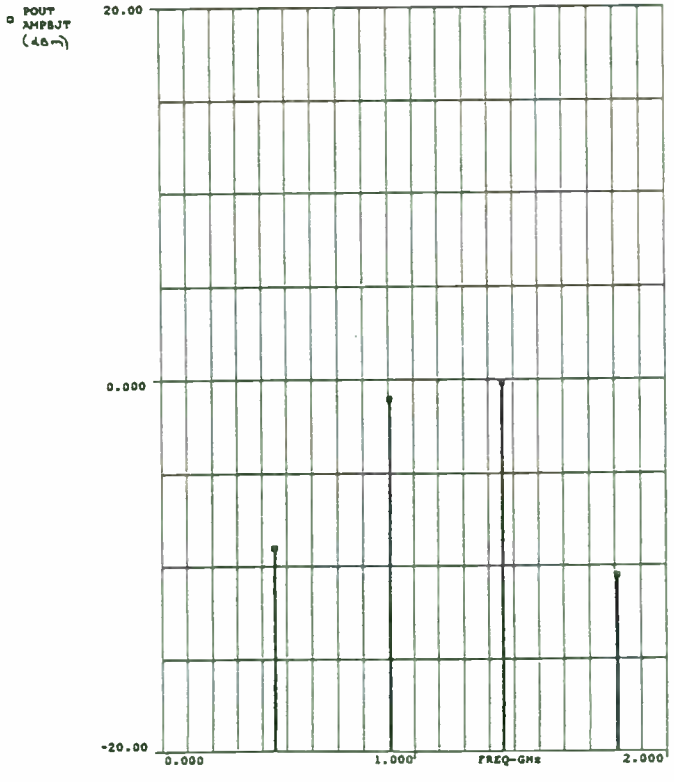


Figure 2: Non-linear Libra™ plot of X3 multiplier

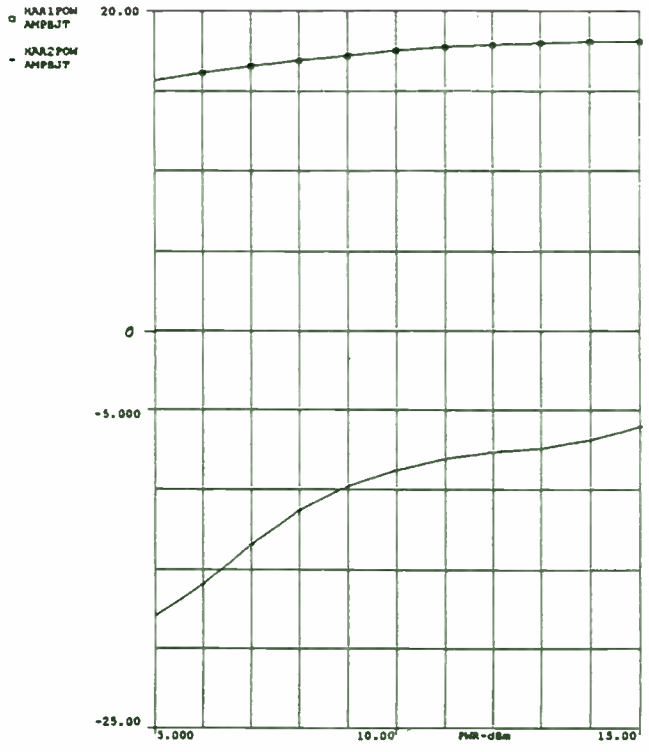


Figure 3: Output Amplier First and Second Harmonics with varying inputs

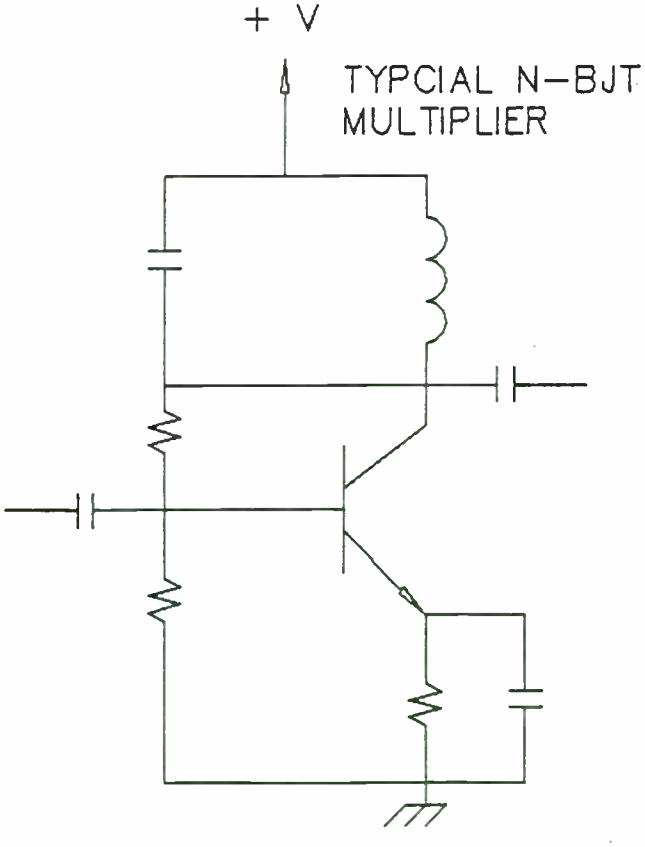
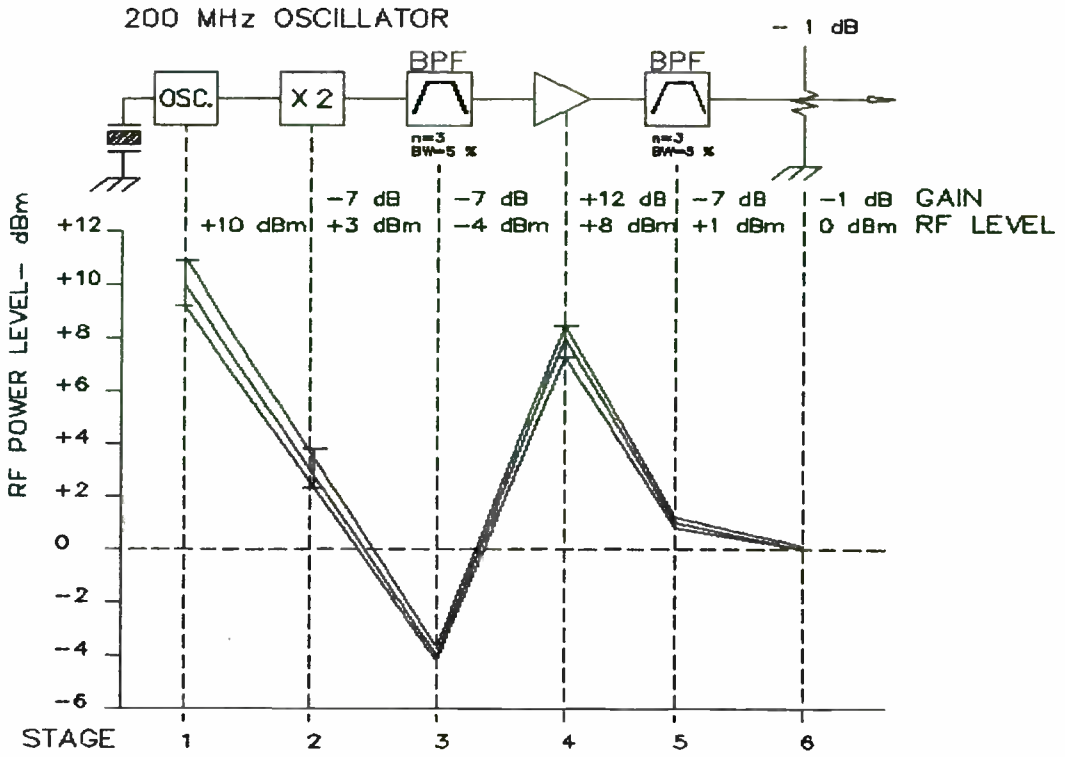


Figure 4: Typical Multiplier Stage



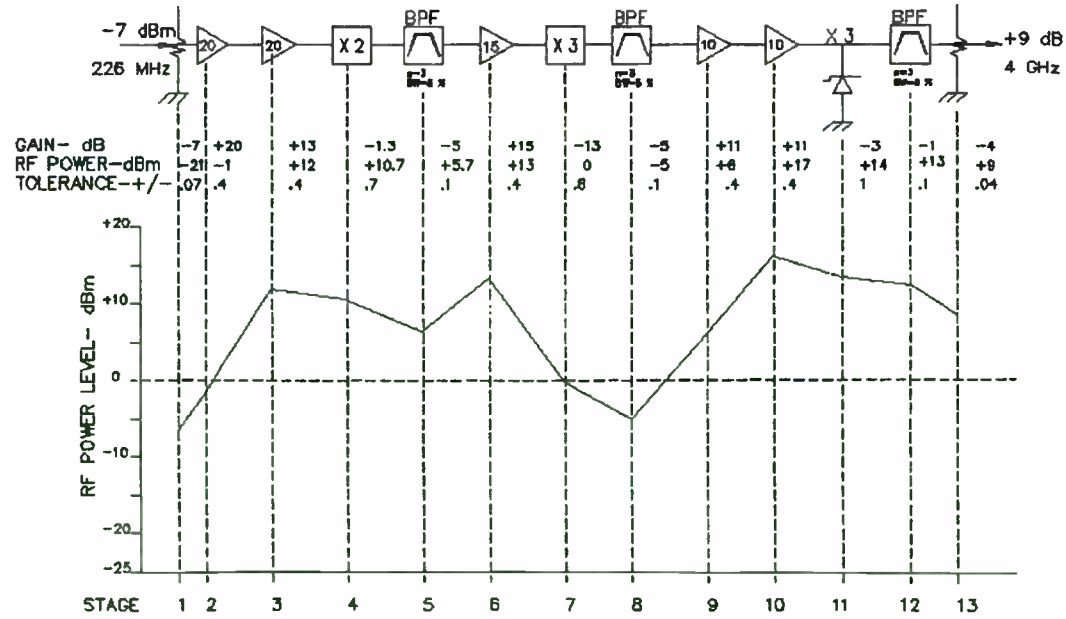


Figure 1: Oscillator and X18 multiplier block diagram

Low Cost Plated Plastic Diplexers For Use In Commercial Mobile Satellite Communications

Chip Scott
Teledyne Electronic Technologies
Microwave Products
1290 Terra Bella Avenue
Mountain View, CA 94043

Global satellite communications are growing at an astounding rate. Never before has the need for rapid and reliable telecommunications been so vital for professional and private success. Over the last several years a tremendous amount of design activity, as well as financial investment, has been placed on the latest generations of light weight, low cost mobile communication systems. As the demands for this market increase we can expect to see substantial pressure placed on cost reduction and enhanced performance.

As with most of today's state-of-the-art communication systems, the proper filtering technology is critical to both cost and performance. In the past, system designers have had basically two options; focus on low cost and sacrifice performance or incorporate the preferred solution, i.e. cavity filters and minimize costs as best as possible. However, until recently the latter option still represented considerable costs due to associated machining processes.

High performance cavity type filters are often used in many communications systems. High performance cavity filters reduce the performance demands of associated system components in areas such as; the noise figure of Low Noise Amplifiers, the gain of antennas and even critical battery life for portable applications. In many of today's commercial applications, cavity type filters are not common, primarily due to their relatively high costs. However, the latest in high performance engineering thermo-plastics can be injected molded and plated to replace the conventional metal housing. The use of plastic significantly reduces the cost of the housing and opens the door to alternative assembly and tuning techniques. An additional benefit of replacing the aluminum housing with plastic is a considerable reduction in weight of approximately 35%.

This product was originally developed to fill a need for a high performance, low cost diplexer for a mobile earth terminal application. Extensive review of the specifications of available engineering thermo-plastics led us to choose a filled Polyetherimide (PEI) material. A block of the candidate material was purchased and machined for evaluation. A duplicate housing was machined out of aluminum for direct comparison purposes. Both filters were silver plated, assembled and tuned in the same manner. Surprisingly, the plastic filter exhibited a more stable electrical performance over temperature than the aluminum housing. Figures 1 and 2, aluminum and plastic filters respectively, show the electrical performance over the temperature range of -35 to +85°C. Note that the plastic filter was tuned slightly narrower than the aluminum filter. Note also that the insertion loss droop of the plastic filter at ambient and hot temperatures is associated with excessively thin plating in this sample unit. The initial success of this test led us to continue development.

The plastics industry in the last decade has seen

an explosion in the development of so called engineering thermo-plastics. Engineering thermo-plastics are plastics used for their performance characteristics and are used as alternatives to glass, metal and wood. Every large plastics manufacturer has a complete line of them. Early on we developed a set of criteria that we used to cull the choices. An important consideration for a microwave filter application was that the plastic must be platable, preferably with silver. Equally important was that the plastic had to be injection moldable. For the filter to perform over a large temperature range we needed a plastic that had a coefficient of linear thermal expansion that was less than aluminum (our usual housing material). This requirement is what led to the choice of a filled material. Fillers in plastics include wood flour, Kaolin (clay), cotton/cloth, mica, and glass. Special care must be taken when specifying a filler since filler can have a considerable impact on the mechanical properties of the molded part (as can color!). Since the original application was for a transmit/receive diplexer the final criterion was the plastic must be dimensionally stable even at high temperatures.

The cost benefit is where the shift to plastic really shines. In quantities of 1000 the cost of a machined diplexer housing is about $30. The equivalent cost of the plastic housing is only $3- even though the plastic that we have chosen is relatively expensive compared to other plastics. The cost of the plating is comparable for the aluminum and plastic housings. The tooling (mold) cost must be considered. The tooling cost for the diplexer is on the order of tens of thousands of dollars. Even so, the non-recurring cost of the mold can be easily amortized over the piece part price on large production runs. The tooling cost can be recovered in part volumes as small as 1000 pieces.

The original diplexer design was an L band diplexer. Due to the skin effect the housing material below about five skin depths is completely arbitrary. The fact that the plastic is a good dielectric has no effect on the electrical performance of the filters. And, since the coefficient of linear thermal expansion is lower than that of aluminum the filters are more stable (less band edge drift) than equivalent aluminum filters.

For high power applications, such as a transmit/receive diplexer there were some concerns about the ability of the plastic parts to dissipate heat. The thermal conductivity of the plastic is only about one one thousandth the thermal conductivity of aluminum. The major concern was the long term and thermal integrity of the silver plating. The original machined and plated parts were subjected to elevated temperatures. The plating was found to adhere without blistering up to about 200°C. At 200°C the plastic undergoes a phase change and the plating blisters. Since the plastic is essentially an insulator the silver plating plays a major role in the dissipation of internally generated heat (due to the insertion loss of the filter). The original machined and plated part was subjected to high power testing. With an insertion loss of 1 dB the plastic part was capable of passing 18 watts with no degradation or damage. The filters are capable of handling higher power levels utilizing specially developed techniques to dissipate the internal heat. An extensive thermal analysis was performed and verified the laboratory results.

An important side effect in the switch from aluminum to plastic is the reduction in weight of the parts by about 35%. The density of the plastic is so low that the weight of the filter is driven by the weight of the aluminum cover. The reduced weight has become a major attraction for manufacturers of portable equipment. One other potential application where reduced weight is important is space (High-Reliability) applications where it costs approximately $30,000 per pound to launch a payload into space. To test the feasibility of utilizing plated plastic filters in space the original plated plastic parts were subjected to a vacuum to test for survivability and passed with no blistering or degradation in performance. The plastic itself exhibits very low outgassing characteristics. The potential use of plastic plated parts in space is still being investigated.

There are still many areas in low cost and plated plastic filters that need to be and are being investigated. One of these areas is to take advantage of the properties of plastic parts and incorporate advanced assembly techniques, such as, snap together parts, ultrasonic welding, and/or solvent bonding. Another area to be investigated is solderless 'tapping' techniques. (A tap point is the electrical connection between the coaxial connectors and the filter input and output

resonators.) And of course, the dream of every filter engineer, automated tuning, which would also go a long way in further reducing filter cost. And finally the low cost of plated plastic filters coupled with higher levels of integration including: low noise amplifiers, power monitors, SWR detectors and even horn antennas can have a significant impact on system performance and reliability while at the same time minimizing costs, size and weight.

Conclusions: Through the utilization of engineering thermo-plastic material and a proprietary plating process, the adhesion and mechanical problems of the past with plated plastic parts have been eliminated. Adding up all the features; light weight, high performance, large volume production and low costs equals an excellent value. With few exceptions this new, *patent pending,* plated plastic technology is the ideal solution for low cost high performance commercial filtering requirements.

TELEDYNE MICROWAVE

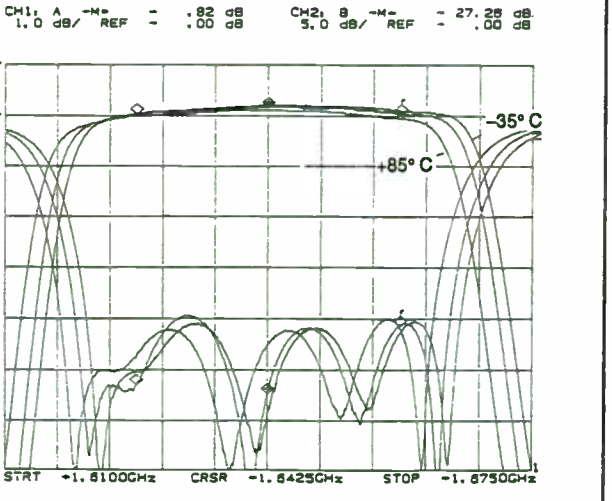


Figure 1

Temperature Performance of Aluminum Diplexer

-35°C to +85° C
6.5 Mhz/Div.

TELEDYNE MICROWAVE

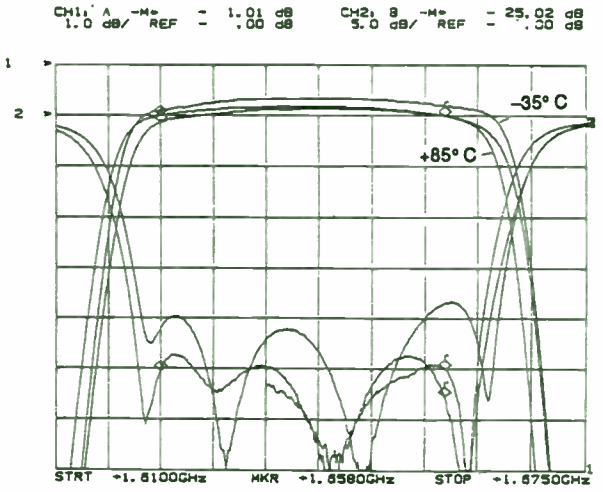


Figure 2

Temperature Performance of Rated Plastic (PEI) Diplexer

-35°C to +85° C
6.5 Mhz/Div.

Satellite Channel Utilization in the Presence of Rain Attenuation

Kaivan Karimi and *Valentine Aalo*
Electrical Engineering Department
Florida Atlantic University
Boca Raton, FL 33431
Phone: (407) 367-3485

ABSTRACT

The most dominant cause of signal degradation in satellite links operating at Ka-band is attenuation due to rain. Presently most rain compensation algorithms are based on the use of a fixed, large fade margin to combat occasional deep fades. However the use of a fixed margin results in an inefficient use of channel capacity for a high percentage of the time. In this paper an adaptive rain fade counter-measure based on the effective utilization of the channel capacity is used for a typical satellite link operating in the Ka-band. Manning's rain attenuation prediction model, based on the rain history of the transmitting and receiving stations is used to determine outage rates both in terms of channel capacity and BER.

I. Introduction

Satellite communications form an essential part of global telecommunication systems, carrying large amounts of data, voice and video, and offer a number of features not readily available with other means of communications. As demand for telecommunication services increases, spectrum congestion forces system planners to use higher frequencies [1]. Currently, most satellite links operate at C-band, but as time passes, there will be increased use of earth-space links operating at Ka-band. The Advanced Communications Technology Satellite (ACTS), which is scheduled to be launched into orbit in 1993, has several experiments on board related to the utilization of earth-space link at Ka-band with downlink frequencies from 19.2 to 20.2 GHz and uplink frequencies from 29.1 to 30.0 Ghz [2]. ACTS mobile terminal program will explore the potentials of Ka-band to meet the needs of future mobile satellite services. The Ka-band, with the promise of less congestion, may also offer increased bandwidth and, therefore, the expansion of system capacity and user services. Also, the accompanying significantly smaller ground terminal equipment are expected to reduce system cost [3]. However, one of the major concerns of the investigators is that satellite channels at Ka-band are subject to severe performance degradations due to propagation effects that occur at this band. For a mobile satellite link, the attenuation factors involved include scintillation, shadowing, multipath fading, and the most dominant factor of all, rain attenuation. In the design of such a link maintaining a fixed,

large fade margin to combat occasional deep rain fades as done in most rain compensation algorithms, results in severe reduction in communication capacity [4]. Traditional fade countermeasures concentrate on the use of power, bandwidth, and site diversity and pay very little attention to the optimum use of channel capacity. In this paper we use the adaptive rain fade counter measure, based on the effective utilization of the channel capacity proposed in [4] to study the performance of a typical ACTS communication link.

II. Rain Attenuation On Satellite Link

Figure 1 illustrates a typical satellite link. Currently, almost all of the satellites in operation have no onboard processing. For a conventional frequency translation satellite, the total system carrier-to-noise density ratio $(C/N_0)_{s,c}$ for clear air (no attenuation) is given by [5]

$$\left(\frac{C}{N_0}\right)_{s,c} = \frac{\left(\frac{C}{N_0}\right)_u \cdot \left(\frac{C}{N_0}\right)_d}{\left(\frac{C}{N_0}\right)_u + \left(\frac{C}{N_0}\right)_d} \qquad (1)$$

where $(C/N_0)_u$ and $(C/N_0)_d$ are the carrier-to-noise density ratio for the uplink and downlink transmissions for clear air, respectively. Note that the quantities in (1) are not in units of dB. Now, let

$$S_u = 10 \log \left(\frac{C}{N_0}\right)_u \quad , \quad S_d = 10 \log \left(\frac{C}{N_0}\right)_d \quad . \qquad (2)$$

Then the total equivalent attenuation due to rainfade for the system (up and down link), is given by

$$A(dB) = \left(\frac{C}{N_0}\right)_{s,c}\Big|_{dB} - \left(\frac{C}{N_0}\right)_{s,att}\Big|_{dB} \qquad (3)$$

where

$$\left(\frac{C}{N_0}\right)_{s,c}\Big|_{dB} = 10 \log\left\{\frac{10^{\frac{S_u}{10}} \cdot 10^{\frac{S_d}{10}}}{10^{\frac{S_u}{10}} + 10^{\frac{S_d}{10}}}\right\} \qquad (4)$$

$$\left(\frac{C}{N_0}\right)_{s,att}\Big|_{dB} = 10 \log\left\{\frac{10^{\frac{S_u - A_u}{10}} \cdot 10^{\frac{S_d - A_d}{10}}}{10^{\frac{S_u - A_u}{10}} + 10^{\frac{S_d - A_d}{10}}}\right\} \qquad (5)$$

In (5), A_u and A_d (in dB) are the uplink and downlink attenuations due to rain, respectively. The over-all system attenuation expressed in dB given in (3) can be shown to be

$$A(dB) = A_u + A_d + 10 \log\left\{\frac{10^{\frac{S_u - A_u}{10}} + 10^{\frac{S_d - A_d}{10}}}{10^{\frac{S_u}{10}} + 10^{\frac{S_d}{10}}}\right\} \qquad (6)$$

It is well known that the attenuations A_u, A_d, and A are approximately log-normally distributed random variables [4], [6]. Therefore, the overall long-term probability density function (pdf) of rain attenuation along the link has the form

$$P_A(a) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln a - m)^2}{2\sigma^2}\right) \quad ; \quad a \geq 0$$

$$P_A(a) = 0 \quad ; \quad a \leq 0 \quad . \qquad (7)$$

The median and standard deviation of A can be determined by anyone of two methods described in [4].

III. Efficient Satellite Channel Utilization

Due to particular constraints of modulation, coding, and throughput, and the absence of onboard processing, in most cases satellite communications system designers implement a fixed link budget margin. For very small aperture terminals (VSAT), this implies insufficient utilization of the channel capacity for a considerably high percentage of the time, especially for a satellite communications system operating at 20/30 GHz band which is subject to severe performance degradations due to rain attenuation [4]. A good performance measure of such a communications system is channel capacity which gives the maximum rate of signal transmission over the channel.

A. Channel Capacity

For a continuous channel with additive white Gaussian noise, Shannon defines the channel capacity in bits per sec. (b/s) by [7]

$$C = B \log_2(1 + \frac{S}{N}) \qquad (8)$$

where B is the channel bandwidth in Hz, and S/N is the signal-to-noise power ratio in the channel. The capacity per unit bandwidth may be written as

$$\frac{C}{B} = \log_2 (1 + \frac{S}{N}) \quad (\frac{b/s}{Hz}) \quad . \qquad (9)$$

Due to propagation fading, the term S/N in (9) is a random variable with arbitrary but known distribution which depends on the characteristics and type of the fading process. This implies that

the channel capacity is also a random variable and therefore imposes performance degradation on the system. Using standard transformation of random variables, the pdf of C/B can be written in terms of that of S/N [4].

In the presence of rain attenuation, the received signal-to-noise power level is given by

$$\left(\frac{S}{N}\right) = \left(\frac{S}{N}\right)_{s,c} - A \qquad (dBW) \tag{10}$$

where $(S/N)_{s,c}$ is the unfaded signal-to-noise power level (in dB) for clear air condition, and A is the total equivalent rain attenuation on the link. The cumulative distribution function (cdf) of channel capacity is given by

$$F_{C/B}(y) = Prob\left\{\frac{C}{B} \le y\right\} = Prob\left\{\frac{S}{N} \le 2^y - 1\right\}$$

$$= Pr\left\{A \ge \left(\frac{S}{N}\right)_{s,c} - 10 \log\left(2^y - 1\right)\right\}$$

$$= Prob\left\{\ln A \ge \ln\left[\left(\frac{S}{N}\right)_{s,c} - 10 \log\left(2^y - 1\right)\right]\right\}$$

$$= Q\left\{\frac{1}{2}\left(\frac{\ln\left[\left(\frac{S}{N}\right)_{s,c} - 10 \log\left(2^y - 1\right)\right] - m}{\sigma}\right)^2\right\} \tag{11}$$

where the variable y = C/B , $(S/N)_{s,c}$, A are all expressed in units of dB, and

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \quad . \tag{12}$$

The long-term cdf of channel capacity in the presence of rain

attenuation, for the various unfaded signal-to-noise ratio values is plotted in Figure 1.

B. BER Degradations

Another measure of the performance of a satellite communication system is the bit error rate (BER). For a given modulation scheme, the system BER is a function of S/N, and in the presence of fading, is also a random variable. By following a procedure similar to the channel capacity analysis presented above, it is possible to determine the cdf of BER performance in the presence of propagation fading. This gives the measure of the availability of the communications link, since it gives the percentage of the time for which the specified BER will be exceeded. For a digital communications system using M-PSK modulation scheme, the probability that any M-ary symbol will be received in error, P_s is given by [7]

$$P_s = \frac{M-1}{M} - \frac{1}{2} erf\left(\sqrt{\frac{E_s}{N_0}} \sin\frac{\pi}{M}\right) - \frac{1}{\sqrt{\pi}} \int_0^{\sqrt{\frac{E_s}{N_0}} \sin\frac{\pi}{M}} e^{-y^2} erf\left(y \cot\frac{\pi}{M}\right) dy. \quad (13)$$

where

$$erf\ (x) = 1 - 2Q(\sqrt{2}x)$$

For $P_s < 10^{-3}$, a useful approximation of (13) is given by [7]

$$P_s \approx 2Q\left[\left(\sin\frac{\pi}{M}\right)\sqrt{\frac{2E_s}{N_0}}\right] \quad (14)$$

where E_s is the energy per M-ary symbol, and N_0 is one-sided power

spectral density. The equivalent energy per data bit E_b is given by

$$E_b = \frac{E_s}{\log_2 M} \tag{15}$$

The relationship between probability of bit-error and signal-to-noise may be denoted by

$$P_b = f\left(\frac{S}{N}\right) \tag{16}$$

The long-term cdf of BER in the presence of overall rain attenuation A is then given by

$$Prob\{ BER > x \} = Prob\left\{\left(\frac{S}{N}\right)_{s,c} - f^{-1}(x) > A\right\}$$

$$= 1 - Q\left(\frac{1}{2}\left(\frac{\ln\left[\left(\frac{S}{N}\right)_{s,c} - 10\log\left(f^{-1}(x)\right)\right] - m}{\sigma}\right)^2\right\} \tag{17}$$

Figure 2 illustrates the long-term cdf of BER in the presence of over-all rain attenuation for a given M-ary modulation scheme. One can see that based on the fading characteristics of the channel, along with relaxation of the BER requirements for applications that are more error tolerant (voice as oppose to data), assuming that the modulation schemes available are limited to uncoded M-PSK, the availability can be improved. Thus during deep fades, the transmission rate may be reduced, or equivalently, the modulation scheme may be changed accordingly.

C. Advanced Communications Technology Satellite (ACTS)

With the launch of the Advanced Communications Technology Satellite (ACTS) into orbit, a new door into the optimum utilization of earth-space link at Ka-band will be opened to us. The Space Communications Technology Center (SCTC) at Florida Atlantic University is one of the several centers with propagation experiments on board ACTS. The rain attenuation prediction model which will be used at the center, is a model based on annualized rainfall statistics developed by R.M. Manning of the Lewis Research Center [6]. According to this model, based on the past rain history of a given location, the link (up or down) attenuation due to rain can be predicted. The exceedance curve for the link between two of the locations involved in ACTS propagation studies (Cleveland, Ohio and West Palm Beach, Florida) using R.M. Manning's model are given in Figure 3. We assume NASA Lewis in Cleveland as the transmitting station (uplink), and FAU-SAT terminal located at Florida Atlantic University's Boca Raton campus (Southern Palm Beach County) as the receiving station (downlink). Table 1 is a typical "Statistics of Rain Attenuation" Table showing the percentage of time for which a certain level of rain attenuation is exceeded for the two links involved. Data for the uplink are taken from past rain history of Cleveland, and similarly data for the downlink are taken from past rain history of West Palm Beach. The last column represents the over-all link attenuation (A) based on an uplink and downlink carrier-to-noise ratio of 45.2 dB and 42.7 dB respectively [2]. The International Station Meteorological

Climate Summary (ISMCS) CD-ROM was used to obtain past rain history of West Palm Beach, and Cleveland. ISMCS is a reliable and accurate source of rain history for over 5800 locations throughout the world, dating back to 1948 [8]. Data from ACTS collected at a receive-only terminal at FAU, will be used to compare our prediction model with the actual levels of attenuation. This will enable us to validate R. M. Manning's Rain Attenuation Model for use in Ka-band.

IV. Summary and Conclusion

Spectrum congestion has forced satellite system planners to consider the use of earth-space links operating at Ka-band. However satellites operating at these high frequencies are subject to severe performance degradations due to the propagating medium. The most dominant factor contributing to signal degradations is the attenuation due to rain. To combat occational deep rain fades, currently most satellite system designers use a fixed large fade margin budget, which reduces the link capacity greatly. In order to utilize the channel more efficiently, an adaptive technique based on the characteristics of the channel and the fading process associated with it needs to be implemented. A long-term statistical analysis of the channel capacity based on the fading statistics of the channel can help the system designers to implement an adaptive modulation scheme, in order to use the channel more efficiently. Figure 2 is the illustration of such analysis. Using the statistical rain data for Cleveland (transmitting station) and West

Palm Beach (receiving station) the cdf of channel capacity in the presence of rain attenuation as a function of unfaded signal-to-noise ratio values is graphed.

Figure 3 illustrates the cdf of BER of several uncoded PSK schemes in the presence of rain attenuation. One can see that for a given unfaded signal-to-noise ratio value, by relaxation of the BER for more error tolerant applications, the capacity of the communications can be increased drastically. For example by lowering the BER from 10^{-6} to 10^{-3}, we can use 16-PSK instead of 8-PSK and still stay under 0.01% of the time for BER exceedance. Thus we can stay with in our BER requirements for over 99.99% of the time.

Using Figure 3, once an acceptable level of BER and modulation scheme are chosen, one can then go back to Figure 2 and choose the appropriate rate of transmission. Using such adaptive technique based on the stochastic characteristics of the fading process of the channel, the utilization of the channel capacity may be done more efficiently.

V. References

[1] J. H. Lodge, "Mobile Satellite Communications Systems: Toward Global Personal Communications," IEEE Communications Mag., Vol. 29, pp. 24-30, November 1991.

[2] D. Chakraborty, F. Davarian, and W. L. Stutzman, "The Ka-

band Propagation Measurement Campaign at JPL," IEEE Antennas and Propag. Mag., Vol. 35, pp. 7-13, February 1993.

[3] B. K. Levitt, "Rain Compensation Algorithm for ACTS Mobile Terminal," IEEE Journal On Sel. Areas In Comm., Vol. 10, pp. 358-363, February 1992.

[4] M. Filip, and E. Vilar, "Optimum Utilization of the Channel Capacity of a Satellite Link in the Presence of Amplitude Scintillations and Rain Attenuation," IEEE Trans. On Comm., Vol. 38, pp. 1958-1965, November 1990.

[5] L. J. Ippolito Jr., *Radiowave Propagation in Satellite Communications*, Van Nostrand Reinhold Co., New York, NY, 1986.

[6] R. M. Manning, "A unified statistical rain attenuation model for communications link fade predictions and optimal stochastic fade control design using a location-dependent rain statistics database," Int. J. Satel. Comm., Vol. 8, pp. 11-30, 1990.

[7] V. K. Bhargava, D. Haccoun, R. Matayas, and P. P. Nuspl, *Digital Communications by Satellite: Modulation, Multiple Access and Coding*, Wiley, New York, NY, 1981.

[8] T. Jarret, "The International Station Meteorological Climate Summary CD-ROM," Bulletin of the American Meteor. Soc., Vol. 73, pp. 1578-1580, October 1992.

Table 1- Statistics of Rain Attenuation on Uplink (Cleveland), Downlink (West Palm Beach), and the Over-all link through ACTS ($S_u = 45.2dB$, $S_d = 42.7dB$).

% time	A_u(dB)	A_d(dB)	A(dB)
0.001	90.29	143.39	141.45
0.002	73.31	115.13	113.19
0.003	64.54	100.60	98.66
0.005	54.61	84.21	82.27
0.010	42.95	65.06	63.13
0.02	33.15	49.08	47.21
0.03	28.19	41.05	39.24
0.05	22.67	32.17	30.50
0.1	16.36	22.12	20.79
0.2	11.24	14.06	13.24
0.3	8.74	10.16	9.70
0.5	6.03	5.95	5.98
1.0	3.05	0.73	1.71

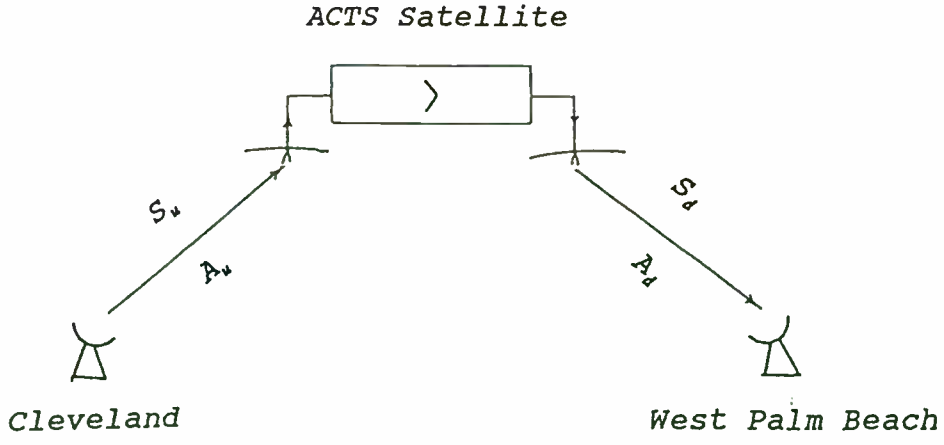


Figure 1. A typical ACTS satellite link.

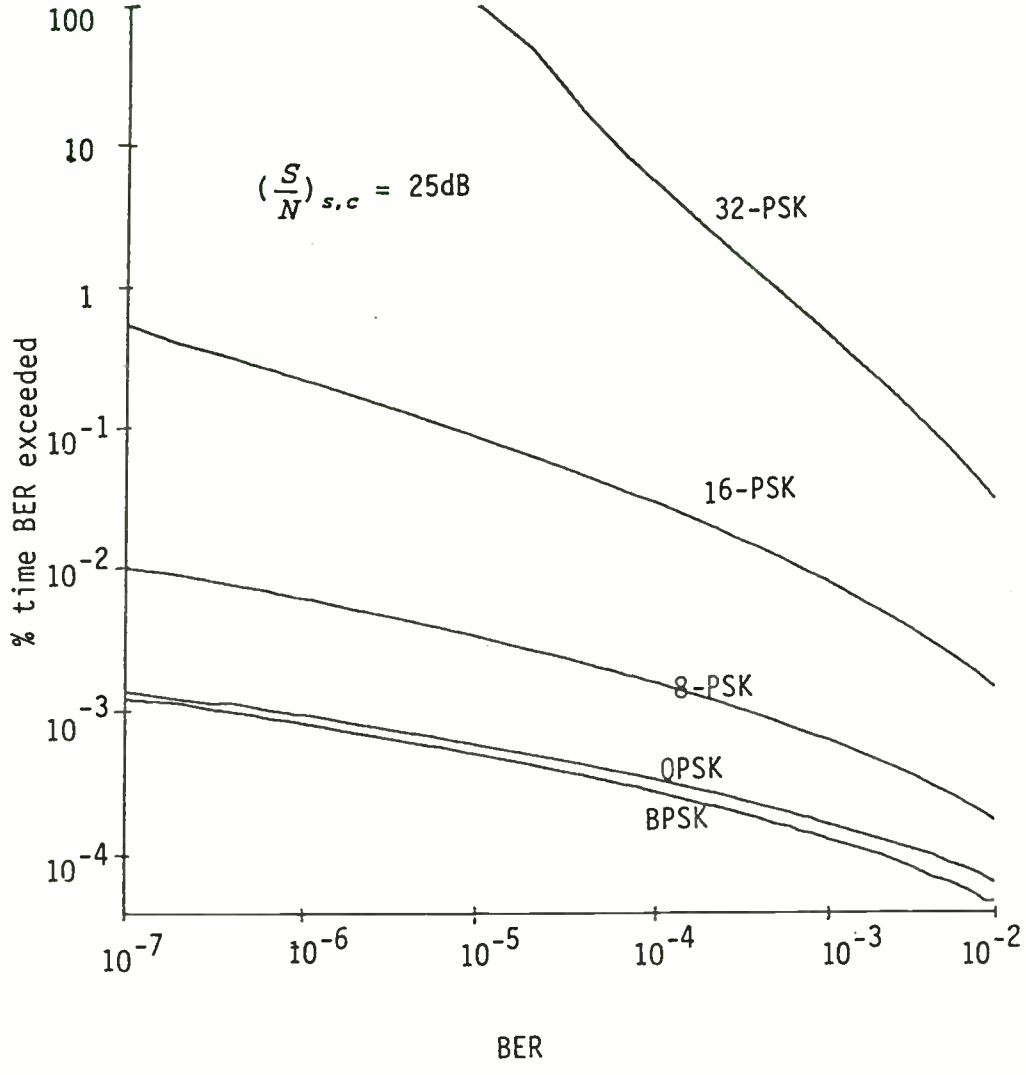


Figure 3. Complementary cdf of BER for uncoded M-PSK scheme in the presence of rain attenuation on the Cleveland - ACTS - W. Palm Beach link.

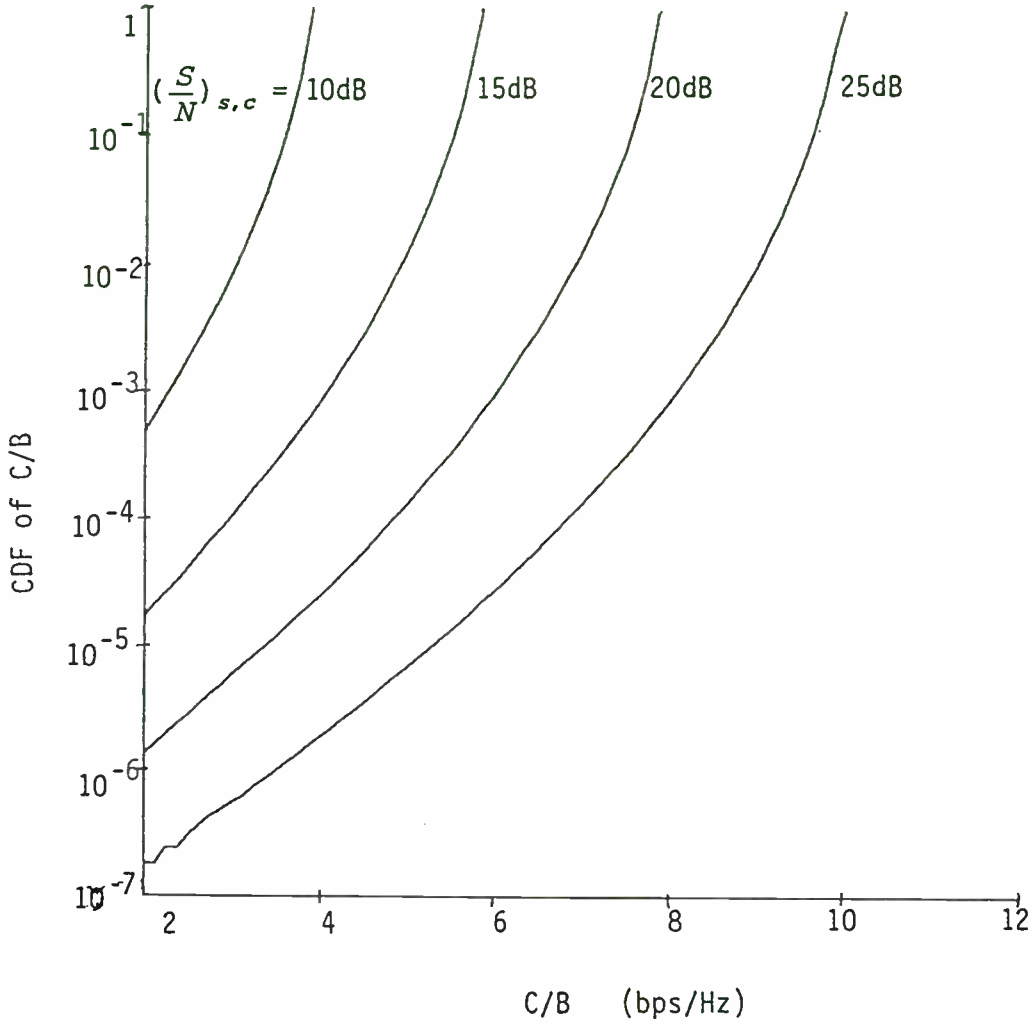


Figure 2. Cumulative distribution of channel capacity in the presence of rain attenuation on the Cleveland - ACTS - W. Palm Beach link.

HARDWARE VERIFICATION OF COMMUNICATION SYSTEM SIMULATIONS

Henry F. Helmken
Florida Atlantic University
500 NW 20th Street
Boca Raton, Florida 33431

Introduction

The Space Communications Technology Center (SCTC), a NASA sponsored Center for the Commercial Development of Space (CCDS), at Florida Atlantic University is developing systems for digital satellite communication. The increasing cost of fabricating hardware prototypes is placing greater emphasis on computer simulations of proposed systems. Signal Processing Workshop (SPW) software from COMDISCO is being used to generate baseband signal files of proposed hardware designs. These files are downloaded to an arbitrary waveform generator which generates real-time baseband signals that can be transmitted over a satellite link. The received satellite signal is down-converted to baseband and digitized for subsequent analysis by SPW. This methodology permits flexibility and speed in system design and performance evaluation.

Mission of SCTC

The Centers for the Commercial Development of Space (CCDS) are non-profit consortia of industry, academia and government that conduct space-based, high technology R & D in areas such as communications, materials processing, bio-technology, remote sensing, automation and robotics, space power and space propulsion. This program was created by NASA in 1985 to maximize U.S. industry leadership in commercial space related activities.

In 1991, a CCDS dedicated to satellite communications was established and headquartered at Florida Atlantic University. Associated with the center are the University of Florida, the University of South Florida, the University of Central Florida and several industrial partners i.e. Harris Corp., Motorola Inc., Cablelabs, Honeywell and Electromagnetic Sciences. The specific mission of this CCDS, the Space Communications Technology Center (SCTC), is to assist U.S. industry in developing the use of digital techniques for transmitting video, audio and data to earth via satellite.

One specific focus of center investigation is the area of modulation and propagation. The goal is to develop propagation models capable of spanning satellite communication links from L-band to Ka-band and use these models for "hardware-in-the-loop" system studies. At Ka band, a series of experiments are planed using the Advanced Communications Technology Satellite (ACTS) satellite. Supporting these experiments will be one of the seven NASA ACTS propagation terminals that has been established at the University of South Florida and a transportable receive only terminal.

System Modeling

Cost of developing prototypes of potential communication systems can be very expensive and time consuming. An alternate approach to an all hardware prototype is to software model, at baseband, the transmitter and receiver characteristics and use general purpose hardware to transmit and receive the modeled baseband signals through a satellite system.

Figure 1 illustrates an implementation of this

process. The Signal Processing Software (SPW) from COMDISCO has been selected for both baseband modeling and analysis software. The current host is a SUN IPX SPARC workstation. Using the block Diagram Editor (BDE) of SPW, a representative transmitting site comprising a data source, an encoding scheme, carrier modulation type, bandpass filtering, Doppler, output power, and transmission plan (i.e. TDMA, CDMA etc.) can be modeled at baseband. Replication of a single transmission site can be used to model a multiple transmitter scenario.

For the selected transmitting scenario, the SPW software is used to generate integer Inphase and Quadrature (I & Q) baseband data files. These baseband waveforms are oversampled, typically 4 to 8 times Nyquist. The I & Q data sets are downloaded via IEEE-488 bus into 2 channels of an Arbitrary Waveform Generator (AWG). The AWG clock is set to produce I & Q baseband signals at the required real time rates. A typical AWG such as the Tektronix 2020 has an internal sampling clock extending to 250 MHz and thus can produce waveforms with a maximum bandwidth of 20 MHz. Each channel has a storage capacity of 256k 12 bit words which can be operated in a continuously looping mode. For example, a CDMA system with a 1 Mhz chip rate, sampled at a 8 Mhz rate, produces a continuously looping 32 millisec baseband I & Q data stream.

Both channels of the AWG are clocked synchronously at the real time rate determined by the bit or chip rate of interest. Each channel is connected to the respective I & Q inputs of a vector modulator which generates the desired complex transmission signal at a specified IF frequency (typically 70 MHz). The IF signal is upconverted and transmitted to the satellite. The received signal undergoes the reverse process i.e. downconverted to IF and vectored demodulated into baseband I & Q components. Each component is sampled via an A/D converter and stored as raw I & Q baseband data for off line analysis.

For the analysis, the data is loaded back into the computer and SPW software is used to model a candidate receiver design which will have to acquire and track the carrier, remove Doppler, and recover the original bit stream. Storing raw I & Q baseband data allows different receiver designs and analysis methods to be applied to the same data set, permitting realistic comparisons (e.g. generate Bit Error Rate (BER) curves) to be made on data acquired under different propagation conditions.

ACTS Satellite

The Advanced Communications Technology Satellite was launched into orbit on August 1, 1993 and achieved operational status on August 1993.

ACTS operates in the Ka band with downlink frequencies from 19.2 to 20.2 GHz and uplink frequencies from 29.1 to 30.0 GHz. These frequency bands are high susceptible to rain fades and hence, for commercial use, robust transmissions schemes will have to be developed. In the Microwave Switch Matrix (MSM) mode, the satellite operates in a wideband "bent pipe" mode and thus suitable for testing a variety of transmission schemes.

ACTS Demonstration Plan

The proposed scheme shown in Figure 1 is now under development and will be tested using ACTS in the MSM mode. Opportunities will also be sought on Low Earth Orbiting (LEO) transponder satellites. Software models of both transmitter and receiver designs and transmission schemes (TDMA, CDMA, etc.) have been developed I & Q baseband data files have been generated by both SPW and special communication software code. These files have been successfully down-loaded to a Tektronix 2020 AWG and coupled to an HP 8782B Vector Modulator and IF performance verified via a Vector demodulator.

The main area of current effort is selecting and testing hardware for an IQ vector demodulator, interfacing I & Q data streams to A/D converters and storing several Gbytes of data at a high sampling rate. The choice for the A/D converter is the Analog

Devices AD872, a 12 bit A/D capable of running at 10 MSPS. For the data recording system, the goal is to have the capability of storing 10 minutes of continuous data at a 10 MBytes/sec sampling rate. The initial storage device will be a video tape recorder, a Panasonic D3 unit, which can accept digital data in NTSC timing format. An alternate high data rate storage device is the Storage Concepts Model 71 RAID system. This latter system employs parallel transfer disk technology and can accommodate 5 GByte storage at continuous transfer rate up to 20 MByte/sec. The probable host for this latter system will be a VME based controller interfaced to a SUN Sparc Workstation. Data will be transferred and stored on 5 Gbyte tape cartridges for subsequent analysis.

Conclusions

As part of the effort to help U.S. industry develop competitive digital satellite system, SCTC is developing a combination software-hardware concept for testing new communication systems. The concept relies on the capabilities of COMDISCO SPW software to model and generate baseband signals of complex transmission scenarios and analyze the received raw baseband I & Q data.

Test baseband I & Q data files have been generated by SPW models and successfully downloaded into an Tektronix 2020 AWG. The AWG has been run at a real time data bit rate of 1 Mb/s, (8 Mhz sampling rate) into a vector modulator generating an IF signal at 70 MHz. This latter signal is available for upconversion to a satellite frequency. The received baseband signal will be vector demodulated and raw baseband data stored via a 12 bit A/D. Data recording will be either via a Panasonic D3 video tape recorder or a Storage Concepts parallel disk array. SPW receiver designs are under development to acquire and track the signal and recover the original bit stream for BER analysis.

References

Advanced Communications Technology (ACTS) Experiments Application Guide. NASA TM-100265 (January, 1990)

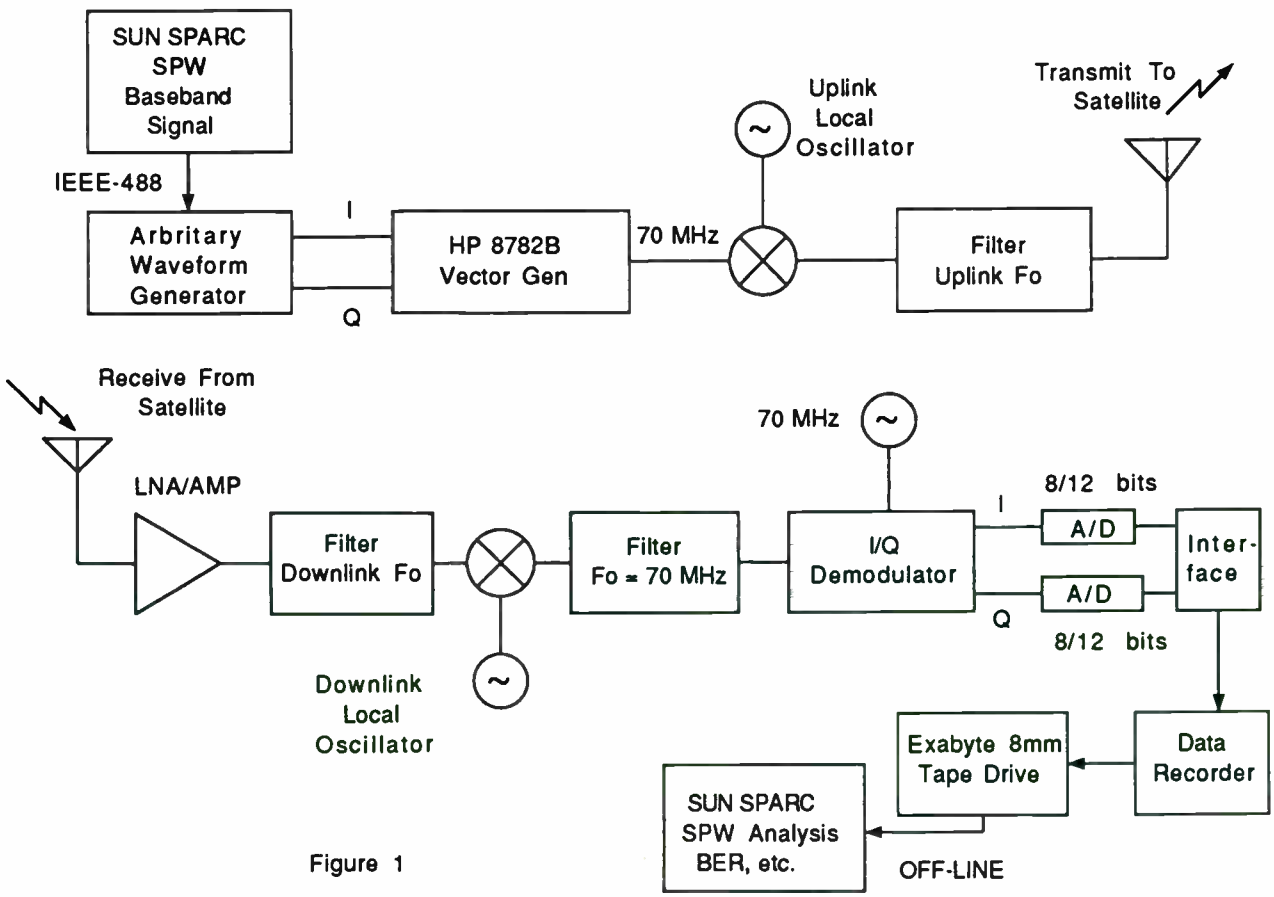


Figure 1

GFO Water Vapor Radiometer

Muhammad A. Malik, P.E.
E-SYSTEMS, Inc., ECI Division
1501 72nd Street North,
P.O. Box 12248
St. Petersburg, FL 33733-2248
Phone: 813-381-2000 ext. 3419
Fax: 813-381-2000 ext. 4810

ABSTRACT

GEOSAT Follow-On (GFO) is the latest Navy altimetry satellite for determining sea surface topography. The required precision is about 3 centimeters in a range measurement of 800 kilometers. Various error sources have to be accounted for to achieve this precision. One error source is the variation of the RF propagation velocity through the atmosphere depending on the moisture content. A total power Ka-band radiometer is provided on-board for atmospheric water vapor delay correction, sharing a broad-band antenna with the Ku-band altimeter.

The radiometer will measure the radiometric brightness temperature of the earth at two frequencies, 22 and 37 GHz, in the approximate footprint of the altimeter return signal. Due to the low level of signal measured, the radiometer receiver design pays special attention to gain stability, and the supply voltage stability. During the radiometer operation, regular calibrations are made using hot and cold sources to maintain the output accuracy. The antenna is required to have very high main beam efficiency and very low sidelobes.

A description of the GFO radiometer and antenna, including the system requirements, correction algorithms and performance projections, is presented in this paper.

INTRODUCTION

Satellite altimetry is a unique tool for understanding and predicting changes in the ocean which have a significant effect on the climate and the life on earth, including the ocean circulation, sea-level and the polar ice sheet volume. There are three major scientific objectives for satellite altimetry [1]:

* Measure the global ocean circulation
* Observe the mean sea-level change; and
* Monitor the polar ice sheet volume

A number of altimeter-carrying satellites have been launched or are being planned as shown in Figure 1.

The current satellite altimeter program, GEOSAT Follow-on (GFO), will provide operational, continuous, global altimeter data on mesoscale sea-surface topographic fronts, eddies, and other oceanographic phenomena to meet U.S. Navy requirements [2].

Under the management of the Space and Naval Warfare Systems Command (SPAWAR) of the Department of the Navy, the GFO program was awarded in 1992 with an anticipated launch in the 1995-1996 time frame, and a specified 8 year mission life.

Ball Corporation, Space and Systems Engineering Division (BSSED), located in Boulder, Colorado, is the prime contractor for GFO, responsible for the payload, spacecraft and launch vehicle, as well as all required ground system modifications. As Ball's team member, E-Systems, ECI Division (ECI), located in St. Petersburg, Florida, is responsible for the Radar Altimeter and other payload systems engineering services. AIL Systems, located in Deer Park, New York, is developing the radiometer hardware and Ball Communications Systems Division, located in Broomfield, Colorado is responsible for the antenna.

GFO PAYLOAD

The GEOSAT Follow On (GFO) Radar Altimeter is a 13.5 GHz, nadir looking, pulse compression radar [3]. It is designed to provide all-weather global monitoring of sea surface wave height, radar cross-section, and range between itself and its nadir point on the sea surface with great precision.

The variability of atmospheric water vapor column abundance produces an unpredictable variation in altimeter range measurements which can be as large as tens of centimeters, and which must be accounted for in order to achieve meaningful results [4]. This variability cannot be accurately dealt with by models. An on-board nadir-looking radiometer is the most viable approach to accurately measure the tropospheric range delay.

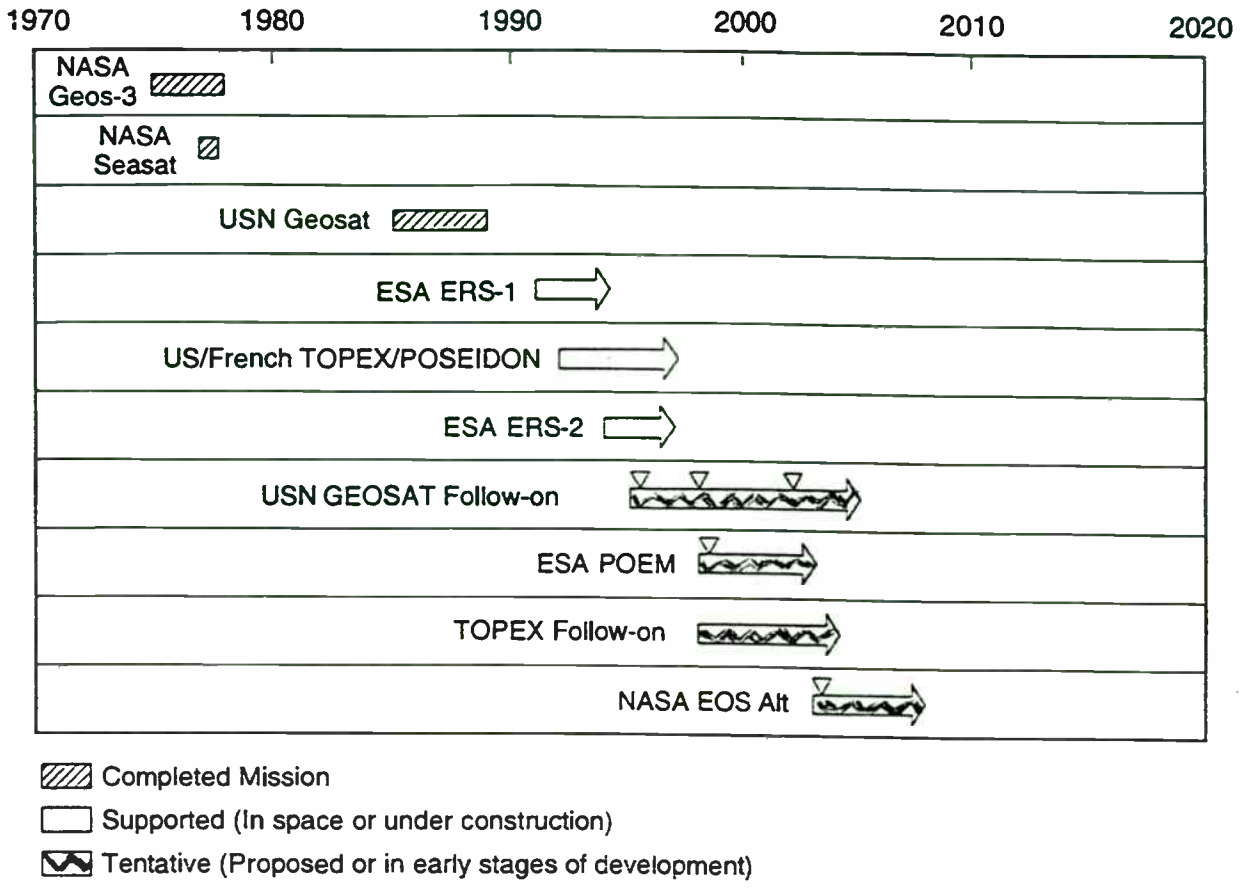


Figure 1. Altimetry Spacecraft: 1970-2020

The GFO Water Vapor Radiometer (WVR) is a two frequency, 22 GHz and 37 GHz, total power radiometer to measure the brightness temperature of the earth. Hot and cold calibrations are regularly performed to maintain the measurement accuracy. The radiometer outputs digital data, which is telemetered to the ground processing center.

The GFO altimeter and radiometer share a tri-band, nadir looking, offset-fed parabolic antenna. This antenna is designed to have 97% main beam efficiency and very low side lobes.

RADIOMETER DESIGN FUNDAMENTALS

The theory of microwave radiometry is described in the literature [5],[6]. In essence, the received power at the radiometer is proportional to the radiation intensity, which is related to the water vapor content. In the microwave region, the Rayleigh-Jeans form of Planck's law can be used to relate the power received to the apparent temperature in the following direct linear form:

$$P = \tfrac{1}{2} A \iint_{4\pi} \tfrac{2k}{\lambda^2} T(\theta,\phi)\Delta f F(\theta,\phi) d\Omega$$

where P = Power received
A = Receiving aperture
k = Boltzmann's constant
λ = Wavelength
T = Apparent temperature
Δf = Bandwidth
F = Normalized radiation pattern
θ = Azimuth angle
ϕ = Elevation angle
dΩ= Differential solid angle

As indicated by the above equation, the radiometric temperature is obtained by measuring the incident power with a well calibrated antenna and receiver.

There are several tradeoffs involved at the system level to satisfy the mission requirements with a minimum size, weight and cost package. These include the frequency selection, radiometer type, number of operating channels, and RF bandwidth.

The radiometric temperature of the atmosphere with respect to frequency, as shown in figure 2 [7], can be exploited for different purposes. The water vapor spectral line centered at 22.2 GHz is suitable for studying properties that depend on the total amount of water vapor present in the atmosphere, and is typically used for radiometry. The liquid water also emits strong radiations. Separate measurements are needed at two frequencies to properly determine the water vapor and the liquid water. For this reason, the frequencies of 22.2 GHz and 37 GHz were chosen.

There are two basic types of radiometers; total power and Dicke, as shown in Figures 3 and 4, respectively. Both of them produce an output voltage proportional to the received signal power. The main difference between them is the Dicke switch which alternates the receiver connection between the antenna and the calibration reference to minimize the time interval available for gain fluctuation.

To appreciate the importance of gain fluctuation, consider a case of antenna temperature of 300 K, a bandwidth of 100 MHz and a measurement interval of one second. For a total power radiometer, the measurement uncertainty (proportional to $1/\sqrt{Bt}$, where B is the signal bandwidth and t is the measurement interval) in this case is 0.03 K or one part in 10^4 of the total noise power. The stability of the radiometer would have to be held to better than one part in 10^4, over a period of one second, so that gain fluctuations do not dominate the sensitivity of the instrument. Historically, Dicke radiometers have been used to circumvent this problem. If the switching time between calibrations is much smaller than the time period of the gain fluctuations, the effect of gain fluctuations becomes almost negligible.

The component technology is advancing, however, such that with a careful circuit design, the receiver gain can be stabilized enough to enable the use of a total power radiometer. The elimination of the Dicke switch results in size and cost reduction. Another consideration is the comparatively higher measurement uncertainty of the Dicke radiometer, which is proportional to $2/\sqrt{Bt}$ because of the time evenly divided between the antenna and the calibration reference. The total power radiometer concept was chosen for GFO because of better measurement accuracy and design economy.

The number of radiometer frequencies or channels is another important consideration. More channels would provide information in different areas of the spectrum about the atmospheric contents but will increase the size and weight also. For the GFO, a two-frequency design was chosen because it satisfied the water content measurement requirements in a compact package.

As shown above, greater RF bandwidth would increase the measurement accuracy, but it can adversely impact the noise figure. Electromagnetic interference (EMI) is another important consideration. For the GFO, the RF bandwidth was specified as 220 MHz for

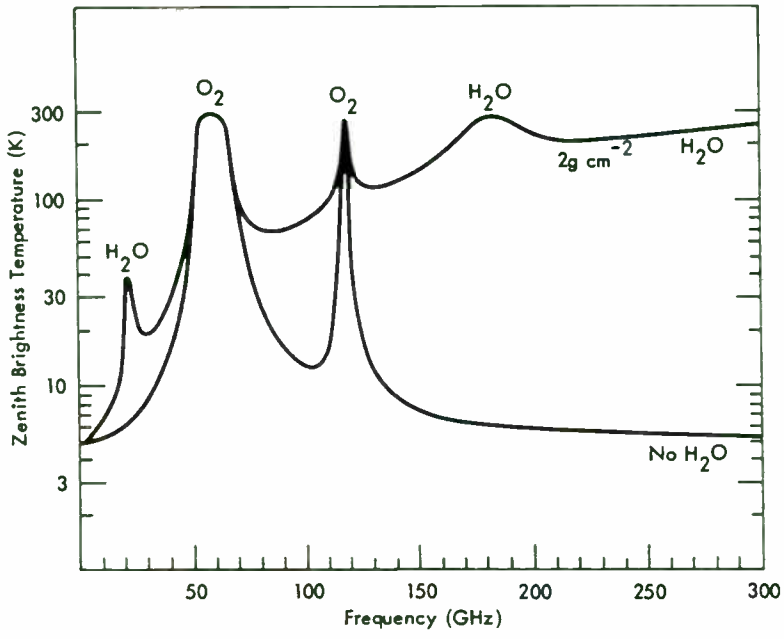


Figure 2. Sky radiometric temperature in the upward zenith direction

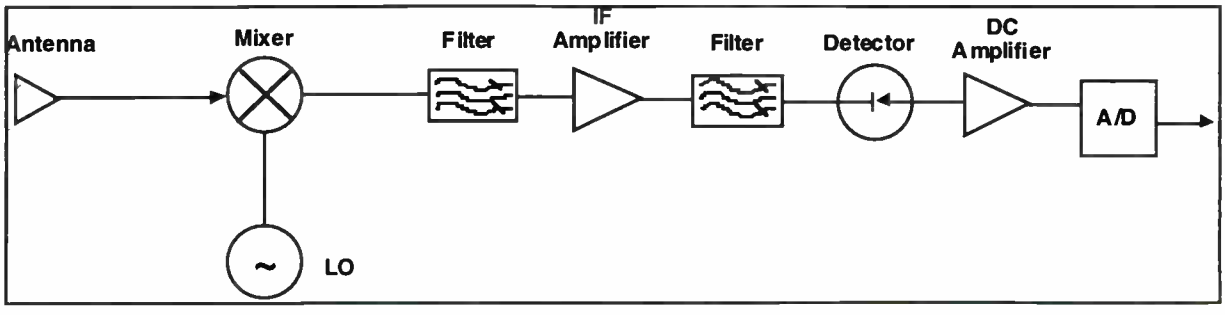


Figure 3. Total Power Radiometer

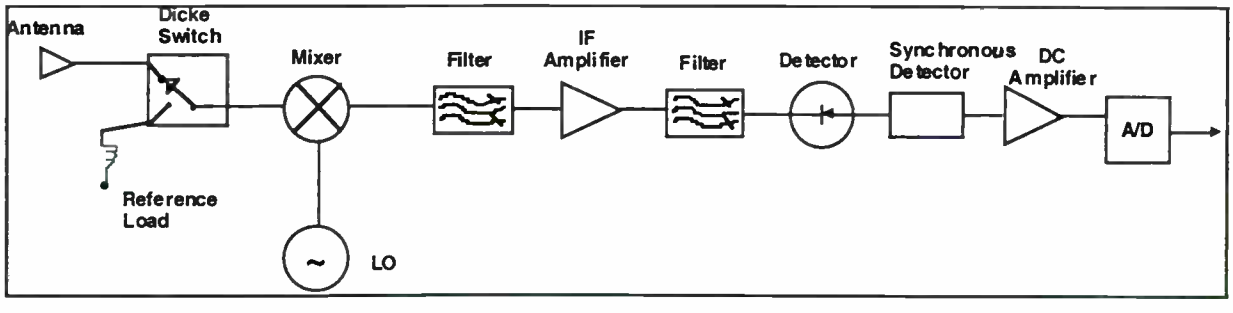


Figure 4. Dicke Radiometer

both the 22 GHz and 37 GHz channels to satisfy the sensitivity and EMI requirements.

RADIOMETER DESCRIPTION

The GFO WVR is a two-frequency, linearly polarized, total power, double sideband, super-heterodyne radiometer. The operating frequencies were selected on the basis of the best path delay error performance and to minimize cost, size and weight. Path delay calculation is based on the emission from atmospheric water vapor which is peaked around 22 GHz. The signal level at 37 GHz is used to indicate liquid water in clouds and rain columns due to microwave absorption and to support the overall water vapor measurement.

The WVR requirements flowdown is shown in Figures 5 and 6. The overall performance requirement is to provide path delay corrections to an accuracy of 1.9 centimeters over the temperature range on-orbit and 8 year life time. These requirements have been related to the system observables and translated into RF subsystem design parameters and component specifications.

WVR Design

The WVR functions to receive the RF energy, downconvert it to baseband, and detect and digitize it in a redundant design for reliability. WVR also includes thermal sensors at selected locations to measure physical temperatures and correct the antenna brightness temperature.

The WVR operating modes are standby, normal and calibrate. It enters the normal state after power-up is complete. Normal mode outputs digital data in a 7-second frame format that includes: brightness temperature measurements at a 2 Hz rate synchronized to the space craft time reference, calibration measurements, physical temperature sensor data, and WVR status and telemetry information. In the calibrate mode, brightness temperature measurements are halted and continuous calibration of the WVR is performed. The WVR can be commanded to the normal or calibrate modes via a data control word from the space craft IAP (integrated avionics processor). All analog outputs are multiplexed and converted to digital data before being output.

WVR Receiver

The WVR receiver, as shown in the block diagram, Figure 7, accepts the RF energy from the antenna at two frequencies, 22 GHz and 37 GHz, each having a prime and a redundant channel for reliability. Switching between the antenna and the two calibration sources is provided by internally latched ferrite switches with greater than 30 dB isolation. The receiver front-end incorporates a double-sideband mixer. The receiver gain fluctuations are minimized by operating well below the device 1 dB compression point and incorporating amplifiers and detectors with low 1/f noise. DC power regulated to 5% level is supplied to the radiometer where it is further regulated to achieve an overall line regulation of 0.1%.

The receiver design includes pre-detection channel bandwidths of 110 MHz at both 22 GHz and 37 GHz. The local oscillators are designed to minimize the frequency drift. The dynamic range is designed to operate over a brightness temperature range of 3 K to 350 K with a system noise temperature due to the receiver electronics of 640 K at 22 GHz and 670 K at 37 GHz. An AGC circuit with a long time constant (minutes) keeps the operating point of the detector constant over long-term gain/temperature changes. Potential frequency interference below waveguide cut-off, such as altimeter frequency of 13.5 GHz, is attenuated to negligible levels before it impacts the RF front-end components. Final filtering occurs in the highly selective low pass IF filter which provides greater than 80 dB attenuation to out-of-band signals. The integration interval is set at 0.5 second, resulting in a 2 Hz output rate. The output voltage is digitized and provided for down link telemetry.

The radiometer weighs 20 pounds, occupies 600 in^3, and dissipates 18 watts.

Cold Horn Assembly

The cold source assembly consists of a 22 GHz and a 37 GHz cold horn, interconnecting waveguide, and redundant thermal sensors that monitor the temperature of each horn and associated waveguide. The cold horns are linearly polarized corrugated horns with approximately 62.5 degree half-power beamwidths. They have maximum sidelobes of -25 dB, a loss of 0.04 dB, and a beam efficiency of 98%. The cold source horns are mounted to provide an unobstructed field of view when looking at deep space, while minimizing the waveguide runs to the receiver front-end.

Antenna

The GFO antenna is a space borne three-frequency, (13.5, 22 and 37 GHz) offset-fed parabolic reflector design that interfaces with the altimeter and the radiometer. The aperture is 41.5 inches diameter and the focal length is 26 inches. The feed is pointed at the central angle of the reflector to minimize spill-over

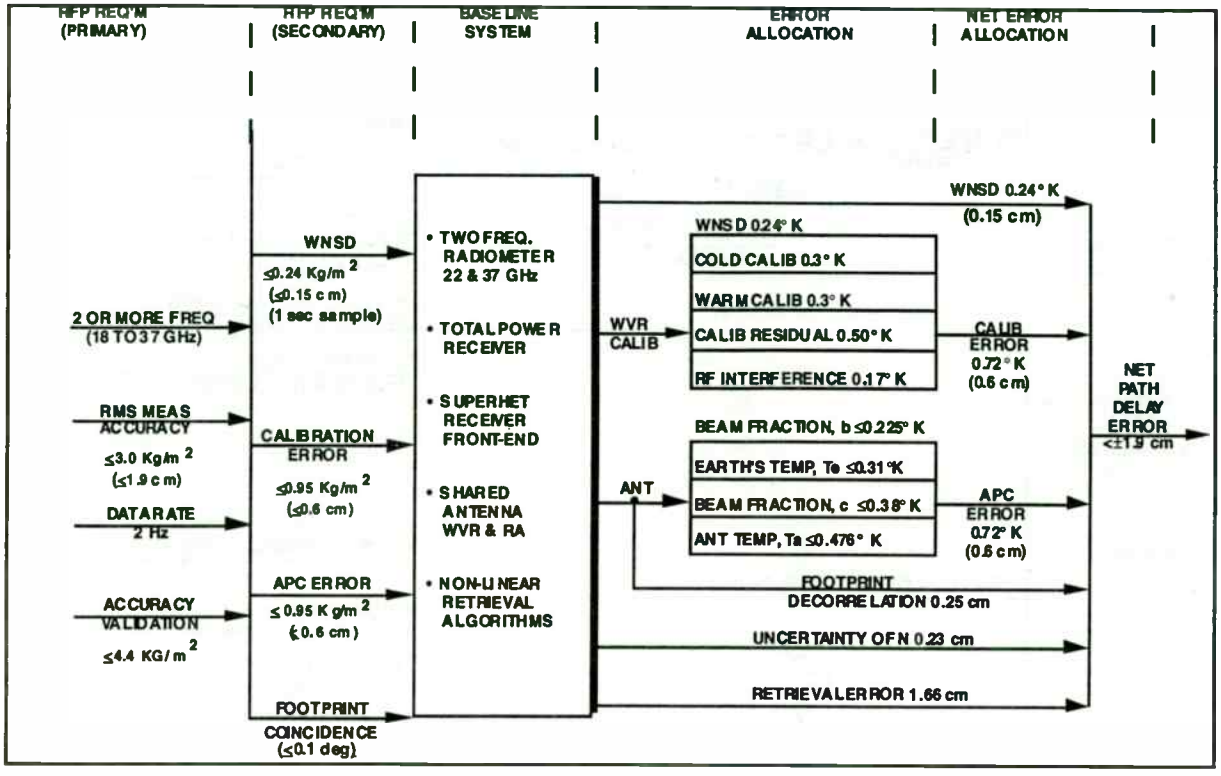


Figure 5. WVR Requirements Flowdown

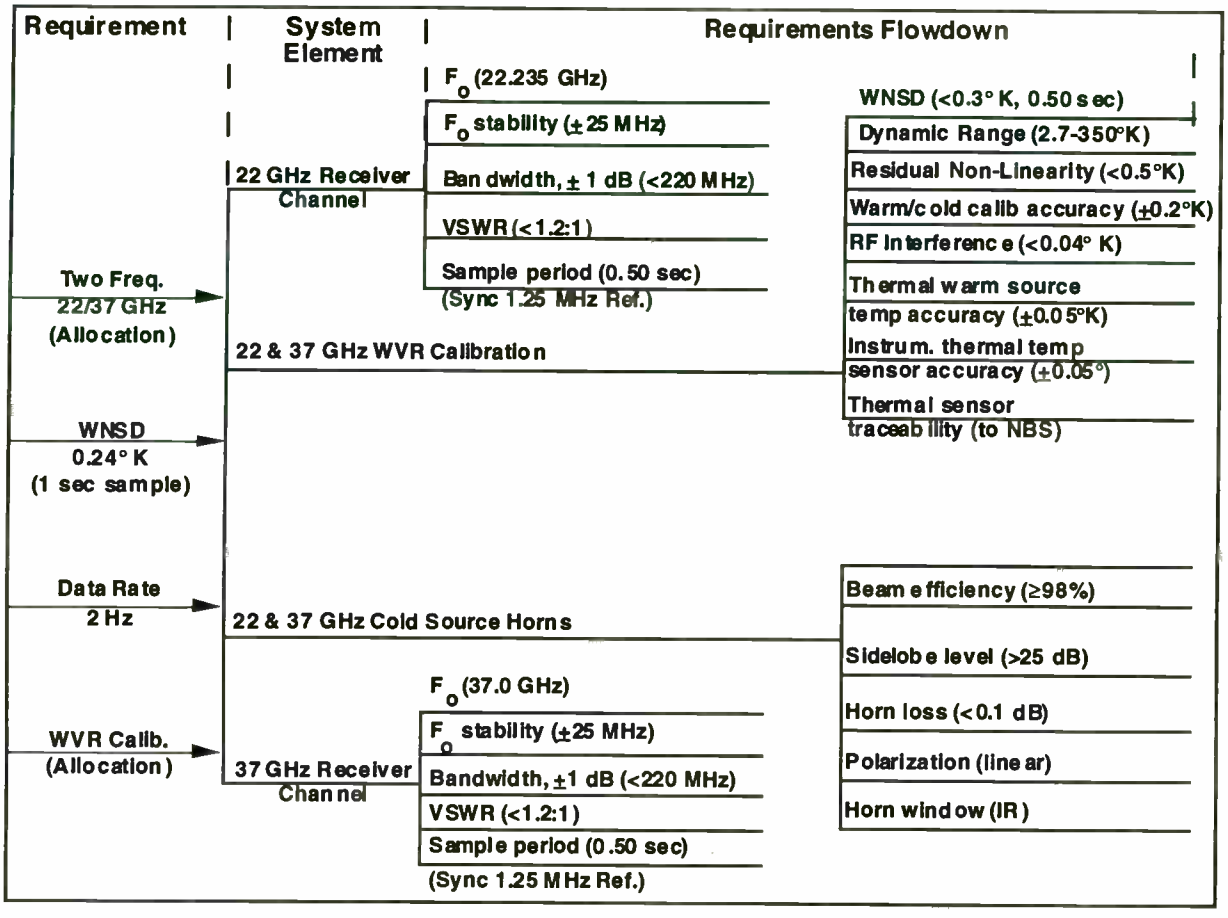


Figure 6. WVR Receiver Requirements/Flowdown

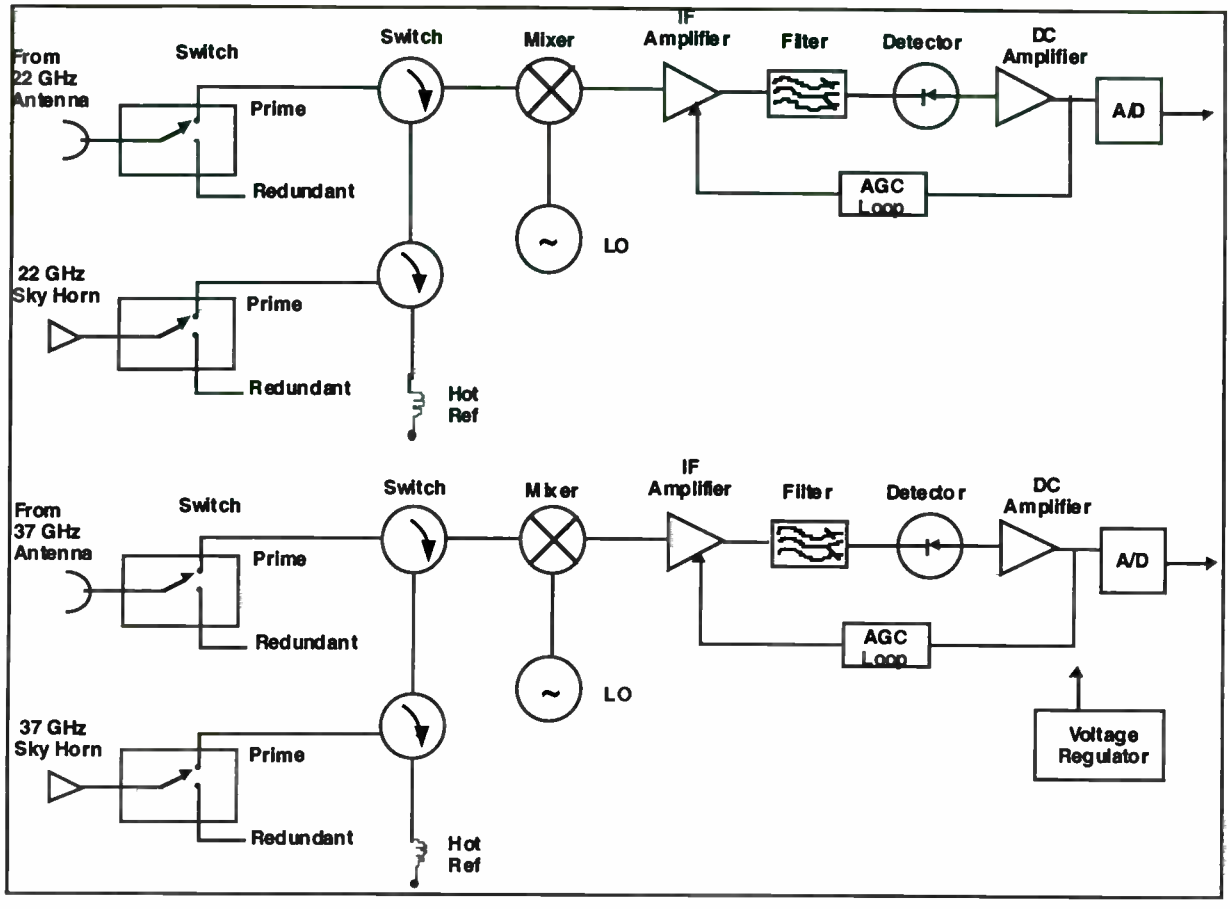


Figure 7. GFO Radiometer Block Diagram

losses. Illumination taper is used to reduce the sidelobes.

The feed is designed for a hybrid HE11 mode. The feed will be located such that the 37 GHz phase center is at the focus of the parabolic reflector. The phase center offset, therefore, exists only at 22 and 13.5 GHz. The antenna requirements are listed in Figure 8. The simulations and test results indicate that these requirements will be met.

WVR PERFORMANCE ANALYSIS

Analyses have been conducted to verify that the radiometer performance shall meet all the specified requirements. The major performance parameters are discussed in this section.

White Noise Standard Deviation

A key GFO WVR performance parameter is the radiometer sensitivity, measured by the output white noise standard deviation (specified as less than 0.15 cm for a one second sample period). This requires a radiometer temperature sensitivity of < 0.24 K. This requirement has been incorporated in the WVR design.

Calibration Accuracy

Radiometer calibration is very important because even small measurement errors, in the range of a tenth of a degree Kelvin, have a significant impact on the wet path delay calculations. For this reason, the radiometer will be thoroughly calibrated on the ground using black body targets and other calibrated instruments. The WVR calibration algorithm accounts for losses, physical temperatures, and re-radiation in the microwave components, as well as non-linearities in the receiver. The on-orbit calibration technique uses a cold source (cosmic space at about 3 K) and a hot source (internal ambient of about 285 K). Both hot and cold calibration measurements are made at uniform 7-second intervals. The net precision of the on-orbit instrument calibration is an rss (root sum square) combination of significant individual error sources, including the white noise standard deviation, cold calibration, hot calibration, calibration residual and RF interference.

Antenna Pattern Correction Accuracy (APC)

Radiation is received by the WVR from antenna sidelobes, reflections from the space craft, and direct spillover into the feed horn. This spurious radiation must be accounted for by the APC algorithm to obtain the mean brightness temperature from the scene averaged over the main-beam solid angle. The net precision of the antenna pattern correction (APC) is an rss combination of the energy received from the regions outside the main beam efficiency region. The APC algorithms, based on the TOPEX microwave radiometer (TMR) heritage [4], are depicted in Figure 9.

Footprint

The altimeter and WVR beam centers are co-boresighted to less than 0.07 degrees. The GFO antenna beamwidths are 1.6 degree for 13.5 GHz, 1.0 degree for 22 GHz, and 0.63 degree for 37 GHz, which provide an effective footprint for the (pulse limited) altimeter of 2 km, and 14 km and 9 km for the WVR. This ensures that the altimeter and WVR footprints are highly correlated and the path delay uncertainty is minimized.

Refractive Index

In space, the energy propagates in a straight line. In the earth's atmosphere, the rays are bent depending on the refractive index of the medium. The determination of refractive index is needed to calculate the difference between the straight line distance and actual distance traveled, or the delay.

Retrieval Algorithms

The retrieval algorithms, relating the measurements to delay calculations, are based on atmospheric models using the radiative transfers and incorporating the satellite radiometry data base. These algorithms represent the results of a significant scientific effort but are still the largest source of uncertainty due to the nature of variables involved.

Path delay Measurement Accuracy

The overall path delay measurement accuracy is an rss combination of the individual error sources including calibration, APC, footprint decorrelation, uncertainty of the refractive index of the air, and the retrieval algorithms. The analyses and test data indicate that the overall radiometer accuracy will satisfy the requirements.

CONCLUSIONS

This paper has presented the requirements and design considerations on the GFO radiometer and antenna. The system level requirements have been allocated to the lower level design specifications and the flight hardware is in development. The GFO design experience is applicable to future space missions.

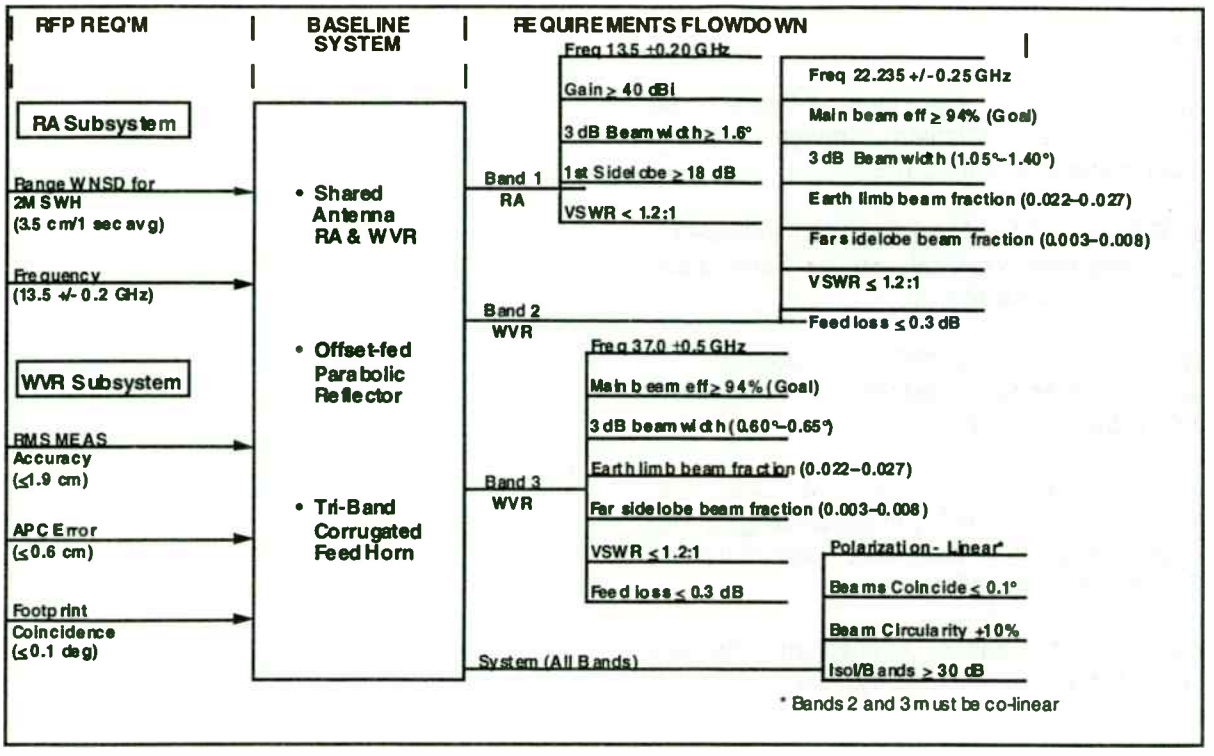


Figure 8. GFO Antenna Requirements Flowdown

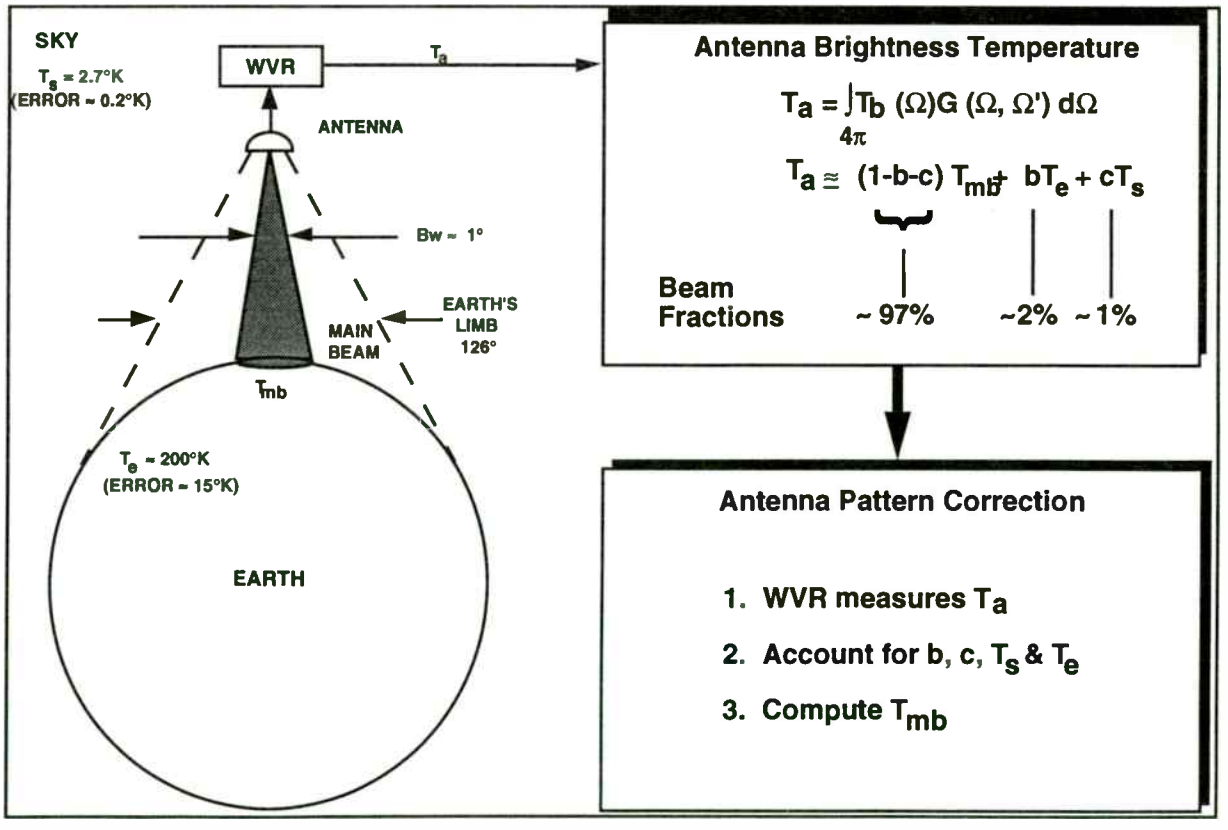


Figure 9. Antenna pattern Correction (APC)

REFERENCES:

[1]. Koblinsky, C. J. et al, " The Future of Space borne Altimetry: Ocean and Climate Change", Joint Oceanographic Institutions Inc., 1992

[2]. Quinn, P.F. et al, "U.S. Navy GEOSAT Follow-on", AIAA Space Programs and Technologies Conference, March 24-27, 1992, Huntsville, AL

[3]. Walker, D. M. et al, " The GEOSAT Follow-on (GFO) Altimeter", AIAA Small Satellites Conference, Salt Lake City, September 1993

[4]. Janssen, M. A., Keihm, S.J., and Ruf, C. S., "The TOPEX Microwave Radiometer", In Proceedings of the Third Specialist Meeting on Microwave Radiometry and Remote Sensing, Boulder, 1992

[5]. Janssen, M. A., editor: Atmospheric Remote Sensing by Microwave Radiometry; John Wiley & Sons, Inc., 1993

[6]. Skolnik, M. I., editor: Radar Handbook, McGraw-Hill Book Company, 1970

[7]. Ulaby, Moore and King: Microwave Remote Sensing, Volume 1, Artech House, 1981

High Power, Low Frequency Microstrip Switches

Sharyn Irons
Eric Higham
M/A-COM MED

Abstract

This paper will describe how low thermal resistance substrates and PIN diodes have been used to produce compact high power switches in the UHF frequency range. These switches, an SPDT and SP3T, are 2.0" x 0.75" x 0.25" and can be used in drop-in or open substrate configurations. Data will be presented for operation at 150W CW input power in the 30-88 MHz frequency range.

Introduction

Switches may perform a variety of different functions in electronic systems. Depending on the system and the desired function, these switches may be located anywhere from the antenna to the receiver or transmitter. The performance parameters required from the switch will be determined, in part, by this location. To best meet this broad range of requirements, switches are fabricated from many different technologies. PIN diode technology offers several advantages to system designers.

PIN diode switches can be faster, smaller and more reliable than electromechanical or ferrite switches. They can handle more power than GaAs FET switches. The need is arising for switches that are small, compact and can handle transmit powers in the UHF communication bands of 100-200 watts.

Mechanical

Switches meeting these requirements do already exist. These conventional electrical and mechanical designs result in large devices. A primary goal of this new design was to minimize package size.

This switch was part of a switch filter in a .062" microstrip multifunction assembly. The package could occupy a 2.00" x .750" area. The electrical connections were .035" x .200" x .002" gold plated copper tabs.

The biggest concern with high power switches is the removal of heat. The major source of this heat is power dissipated in the PIN diodes. The RF design incorporated series diodes to satisfy the size and frequency constraints. With the diodes mounted in series, the thermal resistance path for removal of heat includes the PIN diode and its attachment material, the substrate and its attachment material, and finally the carrier. Low power PIN diode switches are generally manufactured using soft substrates or Alumina (Al_2O_3) for the transmission media. A thermal analysis would show these materials to have a thermal resistance of greater than 200°C/W. This is much higher than what can be tolerated for this application. For high power applications where thermal properties of the substrate material are critical, the substrate selection usually narrows to Beryllium Oxide (BeO) or Aluminum Nitride (AlN). The primary reason for considering these materials is a significant increase in thermal conductivity. Both substrate materials will satisfy the thermal requirements, but AlN offers some additional advantages.

Some of the advantages of AlN are higher and more stable thermal conductivity, lower coefficients of thermal expansion (CTE) and non-toxicity. The thermal conductivity values range from 70-320 W/m°K. This property remains more stable than BeO over normal operating temperatures, as shown below[2].

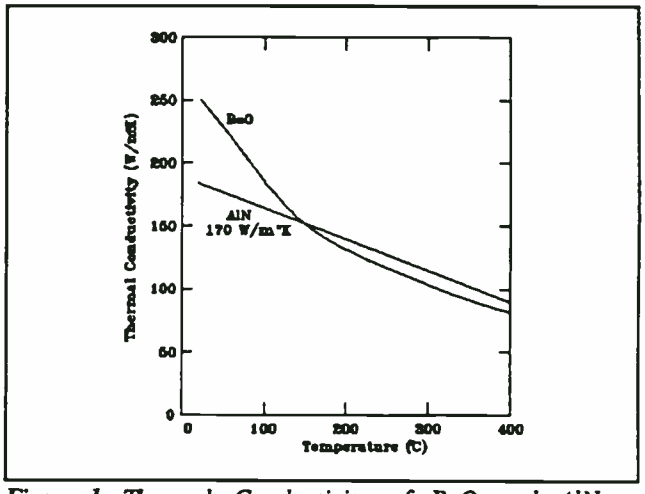


Figure 1 Thermal Conductivity of BeO and AlN vs Temperature

The CTE of AlN (4.5 ppm/°C) is a better match than BeO (8 ppm/°C) to Si (3.5 ppm/°C). Since Silicon PIN diodes and driver IC chips will be used in the switches, the likeli-

hood of cracking, caused by unequal expansion rates is greatly reduced.

The final component in the thermal resistance path is the carrier. The most significant factor in the selection of the carrier material is high thermal conductivity. This is important for transporting heat away from the PIN diode during high power operation. Another very important thermal property to consider is Coefficient of Thermal Expansion. Large CTE mismatches between substrate and carrier can cause damage when subjected to temperature excursions. Solder attachment, for example, will mechanically restrict displacement, and force the materials to bow upon cooling. This bowing can create an uneven ground plane, degrading RF performance, or cause cracking throughout the ceramic.

For this design, Molybdenum was chosen for the carrier material. Thermal properties of some other possible materials are listed below. Kovar and Aluminum have serious deficiencies and cannot be used in this application.

Material	CTE (ppm/°c)	Thermal Conductivity (W/m°K)
W/Cu	6.5	190
MCX-622™	6.2	170
Mo	5.0	140
Kovar®	5.9	14
Al	23.0	171
AlN	4.5	170

To minimize the combined component to substrate to carrier thermal resistance path, the attachment method was chosen to be Au80/Sn20 solder. Other lower temperature solders were considered but discarded to avoid solder reflow during installation. Electrical connections are made with Sn62 or Sn63.

The materials in the thermal resistance path have been chosen in order to insure proper heat transfer from the PIN diodes. AlN was selected as the substrate material, based on the thermal properties. Processing techniques need to be investigated.

AlN ceramic is a mixture of AlN powder, organic additives and sintering agents. The composition of the mixture can be varied to result in a range of practical thermal conductivities from 70 W/m°K to 270 W/m°K. The highest AlN thermal conductivity, 320 W/m°K [3], is a theoretical value which could be obtained only if there were no impurities in the process. However, the sintering process is more difficult for "pure" AlN because it does not densify well. For this application, a thermal conductivity of 170W/m°K was selected.

The chemical structure of a nitride, such as AlN, differs from that of common oxides (Al2O3, BeO). Fewer grain boundaries in high thermal conductivity AlN inhibit good adhesion of thin film metallization. Thick film processing does not completely eliminate concern, either. Thick film pastes contain glass particles which are not CTE compatible with AlN. This mismatch causes cracking of the pastes upon drying. Development of thick film pastes which improve adhesion has been reported [4], but they have not been widely used. In either case, the adhesion may also be affected by processing. Chemical processing agents can cause a reaction which alters the chemical structure of the surface of the AlN substrate. Laser processing can also cause chemical reactions which will disassociate the AlN substrate into Aluminum around the lasered area. This will create a conductive path through the substrate.

Observing these precautions can simplify the processing of AlN. This substrate was manufactured with laser drilled vias and square cut-outs. The circuit was sputtered with nichrome, nickel (1500 angstroms avg) and gold (3000 angstroms avg), using thinfilm technology. To improve solderability and bondability, 100 μin. of gold was electroplated onto the circuit. A final layer of silicon dioxide was deposited for handling protection and insulation between the driver lines with high voltage potential.

The final switch assembly can be used in open substrate or packaged form. A cover provided protection for the bond wires and components and allowed the device to meet the gross leak requirements. Because of space constraints, the cover rested on the substrate. This dictated the cover material to be non-metallic. A plastic with a low CTE would have been ideal, but the CTE of plastic is high. Because of the CTE mismatch, a flexible epoxy seal was used to meet gross leak requirements. Prototype units incorporated ME7155-AN, stress-free epoxy with a Torlon® 4203 cover. Production units used a glass-fibre filled plastic cover, with a silastic (silicone-rubber type) epoxy.

Electrical

To minimize space for the required bandwidth, an all series diode design was selected. The switches required only 37 dB isolation. This is achievable with two series diodes. The all series RF design resulted in bias current from a single supply. The PIN diodes were oriented with cathode to ground at the common junction. This allowed the forward bias current to be generated by the positive supply. Since this supply is much smaller in magnitude (+5V versus -100V), the DC dissipation of the switch is minimized.

Having decided on the topology, a diode was selected to meet these demands. An important consideration for a high power switch is heat dissipation. In the all series topology, the diode, substrate and attachment methods all contribute to the thermal resistance path. If this path is not low enough, the PIN diode junction will reach too high a temperature and fail.

The thermal resistance path from the diode (80Au/20Sn solder, Aluminum Nitride substrate and carrier) has a calculated thermal impedance of 4°C/W. To determine the maximum allowable diode thermal resistance, the expected power dissipation and maximum junction temperature rise must be determined.

The device had to survive a load VSWR of 4.5:1.

This, coupled with a 1.25:1 maximum VSWR for the switch, results in a standing wave pattern with maximum power of:

$$P_{max} = P_{inc} \cdot [2VSWR/VSWR+1]^2$$
$$= 150W \cdot [(2 \cdot 5.625)/(5.625+1)]^2$$
$$= 432W$$

This worst case power represents a peak RF voltage of ≈ 208V. Since -100V back bias voltage is used, the breakdown voltage of the PIN diode must be specified as greater than 600V. PIN diodes with this breakdown voltage routinely have forward resistances of 1 Ω or less at forward bias levels of 100 mA.

For a series mounted diode:

$$P_d = 200R/(100+R)^2 \cdot P_{max}$$

R = Diode resistance.

For a 1 Ω diode with P_{max} = 432W,

$$P_d = 8.47W$$

For this particular requirement, the maximum operating temperature is specified as 100°C. The maximum junction temperature of PIN diodes is conservatively specified at 175°C. This allows a 75°C junction temperature increase.

$$\Theta_{total} = 75°C/8.47W \approx 9°C/W$$

With the remainder of the thermal resistance path being 4°C/W, the PIN diode must be less than 5°C/W for the design to accommodate the worst case power conditions.

For this device to be useful in communications applications, harmonic and intermodulation performance also needs to be considered. It has been shown[1] that, at low frequency, the intercept point of a switch can be calculated as:

$$IP_2 = 34 + 20 \cdot \log(F \cdot I_O \cdot \tau / R_s)$$

where: IP_2 = Second order intercept point
F = Frequency (MHz)
I_O = Bias Current (mA)
τ = Lifetime (μsec)
R_s = Diode resistance

The requirement was for harmonics less than 65 dBc. Second order intermodulation products are 6 dB higher than a second order harmonic, so this implies:

$$IP_2 = \text{Intermod Level(dBc)} + P_{inc}$$
$$= 71 \text{ dBc} + 51.7 \text{ dBm}$$
$$= 122.7 \text{ dBm}$$

Using the quantities we have discussed (R_s = 1Ω, I_O = 100 mA, F = 30 MHz, IP^2 = 122.7 dBm), the equation for IP^2 can be solved to give:

$$\tau_{min} = 9.1 \mu sec$$

Once the desired switch performance was related to the diode parameters, M/A-COM's Burlington Semiconductor Operation was able to supply a CERMACHIP™ PIN diode to meet the following requirements:

$\Theta \leq$ 5°C/W @100 mA
$\tau \geq$ 10 μsec @100 mA
$C_j \leq$ 0.5 pF @-100V
$R_s \leq$ 1.0Ω @100 mA @100 MHz
$V_b \geq$ 600V

The RF circuit is composed of transmission line with 060" x .060" pads to mount the diodes on. The bias is injected and returned for each arm through discrete inductor chips. The DC blocking and RF bypass capacitors are ceramic parallel plate chips.

DC blocking and bias network components (RF bypass capacitors and inductors) will define a filter structure. Ordinarily, the values of these components are selected to result in a filter with a lower cut-off extending well below the lowest frequency of operation. This ensures low insertion loss and VSWR in the operating band, but results in large values of capacitance and inductance for operation in the 30-88 MHz band. As inductance value increases, the self resonant frequency (SRF) decreases. For the device to behave as predicted, the SRF must be above the highest frequency of operation

Designing the filter to have a cut-off frequency below 30 MHz minimized component value, and physical size. The reactances of the components in an "on" arm were modeled and optimized. The final component values were selected to be industry standard values with SRFs above 88 MHz.

The schematic and modeled results are presented below.

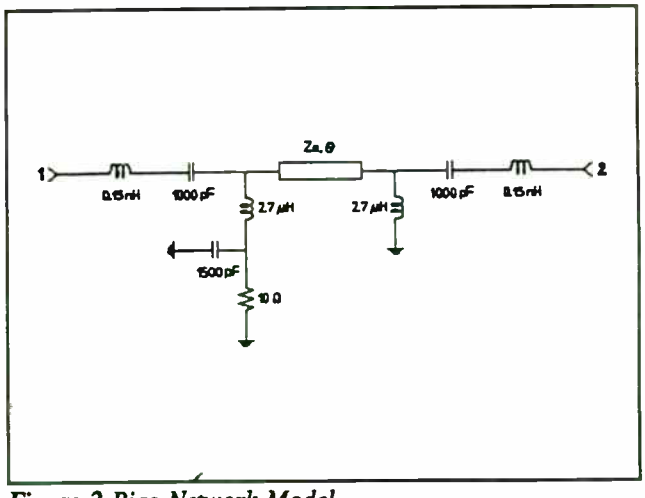


Figure 2 Bias Network Model

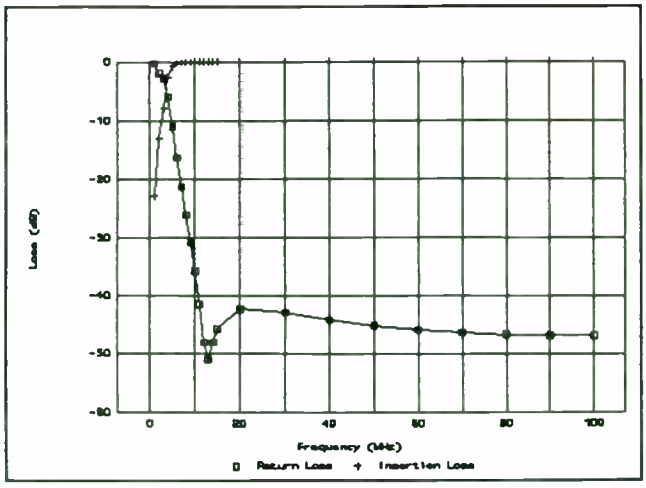


Figure 3 Filter Response of Bias Network

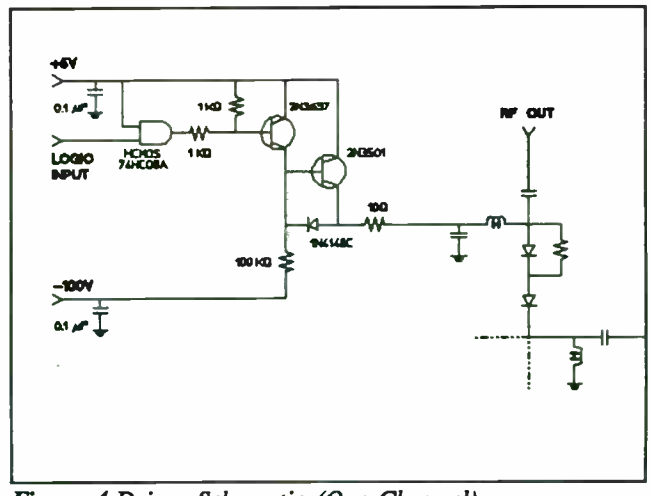


Figure 4 Driver Schematic (One Channel)

Driver

The driver, shown schematically with the RF section below, was designed to provide forward current of 100 mA for insertion loss and reverse voltage of -100V for isolation. The slow switching speed (50 µsec) requirement allowed a design which minimized driver components.

There is a potential drawback to this approach. A provision must be made to back bias the common junction diode. If the reverse bias signal is applied at the anode of the second diode, the first diode will see little or no back bias. The incident RF charge will be high enough to cause the junction diode to compress, or change state.

This problem is solved by the addition of a bypass resistor around the second RF diode. Under reverse bias conditions, there is very little current flow, and the -100V bias is applied to the diode at the junction. Of the two off arm diodes, the junction diode will have the highest incident power, and therefore needs the highest back bias voltage. The second diode is isolated from the input power and needs little or no back bias.

During forward bias, the PIN diode becomes a very low resistance ($\approx 1\Omega$). The bypass resistor is chosen to be large enough so that the parallel resistance of the network is essentially the PIN diode resistance. Over the 30-88 MHz frequency range, this technique causes very little impact to the RF performance.

Conclusion

A design has been outlined for low frequency, high power switches. These design techniques, coupled with Aluminum Nitride substrate and processing technology, and CERMACHIP™ diode technology have enabled the development of small, high power switches that can be configured to meet a variety of packaging requirements. Present devices have a plastic cover which allows the seal to meet gross leak requirements. This design can be used as a building block for drop-in, connectorized or surface mount applications. The design, processing and packaging techniques presented here offer a smaller, more reliable solid-state approach to lower frequency switches.

The RF performance of the completed device is summarized below:

Requirement	Performance	Comments
Input power:	150W CW	4.5:1 Load VSWR
Switching Time :	34 µsec	50% HCMOS to 10% RF
	1.5 µsec	50% HCMOS to 90% RF
VSWR :	1.25:1	
Harmonics :	-70 dBc	
Intermod Distortion :	-25 dBc	2nd
	-22 dBc	3rd
	-30 dBc	5th
	-45 dBc	7th
Isolation :	-40 dB	
Noise Floor:	-155 dBm/Hz	
DC Supply :	+5Vdc @ 200 mA	
	-100Vdc @ 5 mA	
Temperature:	-40°C to +100°C	

The finished switches are shown below.

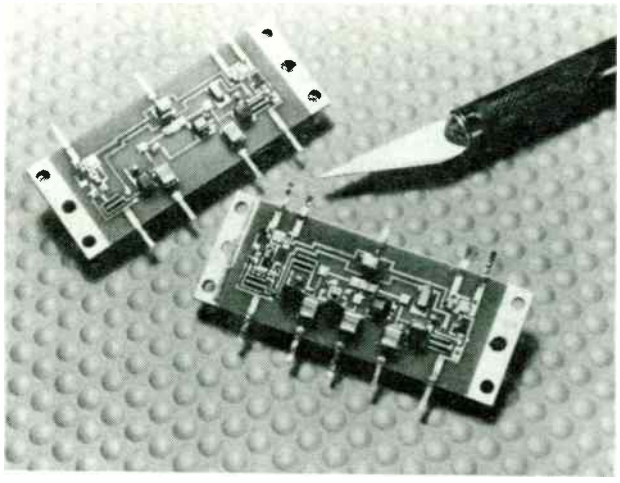


Figure 5

References

[1] R. Caverly and G. Hiller, "Predict PIN-diode switch distortion," *Microwaves & RF*, p. 8, January, 1986.

[2] J. B. Blum and K. Anzai, "Aluminum Nitride Substrates for Hybrid Microelectronic Applications," CeraTronics Application Note.

[3] G. A. Slack, "Nonmetellic Crystals with High Thermal Conductivity", *J. Phys. Chem Solids*, vol. 34, 1973, 321-335.

[4] Yamamoto *et.al*, *39th Electron. Comp. Proceedings*, pp. 23-28, May 1989.

FMCW RADAR ARCHITECTURE

K. V. PUGLIA
M/A-COM, INC.
MICROELECTRONICS DIVISION
LOWELL, MA 01853-3295

1.0 INTRODUCTION

The paper presents the results of a program to develop a low cost, general purpose FMCW radar set. The program was conducted to investigate and to determine those elements of system design as they relate to RF components and subsystems and to demonstrate the hardware effectiveness as a more general purpose range measurement instrument.

The general purpose nature of the hardware is emphasized as applications are limited only by the imagination and awareness of potential users who are confronted with a problem. In addition to obstacle detection, the hardware is suitable for stand alone ranging or positioning applications, traffic control and monitoring, and perimeter security or surveillance. The various applications would require changes to the processor algorithm (software); the algorithm being unique to each application.

The system architecture and many of the system parameters were based upon an earlier M/A Technical Memorandum by M. Hines (reference 1). In the development cycle, it was decided to employ a digital signal processor so that a more general purpose device could be constructed.

The program objectives included the manufacture of an engineering model with the attendant requirements of a microwave assembly, operator display, and packaging and mounting considerations.

The program was conducted over a one year period and included the following tasks:

1. system architecture development
2. component development
3. prototype manufacture
4. algorithm development
5. operational laboratory and field tests.

The tasks are reviewed and results are presented with respect to those tasks.

A summary is presented which delineates the more significant events of the program and offers some suggestions for future study or hardware feature improvement.

2.0 THEORETICAL CONSIDERATIONS

The discussion of the theoretical aspects FMCW radar is quite elementary and is offered only to acquaint the reader with the basic principles necessary to understand the hardware implementation. The interested reader is referred to reference 2 for a more comprehensive presentation.

2.1 FMCW Radar Operation

FMCW radar are amongst the first types employed in low cost ranging applications. Their low cost was primarily the result of a simplified architecture which eliminated some of the components required in pulsed type ranging radar; specifically, the local oscillator, automatic frequency control, etc. The FMCW radar may be visualized as the electronic dual of the pulsed type radar where the timing mark is changing frequency as contrasted with changing amplitude (PAM). The question of range resolution is, nonetheless, determined by the width of the transmitted frequency spectrum; a wider transmitter spectrum resulting in better range resolution.

For the elementary FMCW radar depicted in Figure 2.1, the frequency at the output of the mixer is related to range of a fixed target, i.e. zero relative velocity, in accordance with the formula:

$$f_S = 2R\ (\Delta F/\Delta t)/c\ ,$$

where,

f_S is the output frequency,
R is the target range,
$\Delta F/\Delta t$ is the transmitter frequency rate of change, and
c is the speed of light.

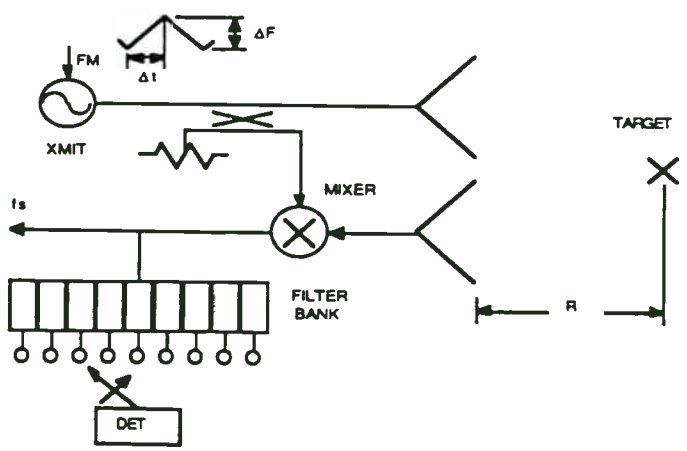


Figure 2.1 Simple FMCW Radar

A filter bank , corresponding to respective range bins, and detector is employed to sequentially sample the energy within each range bin; if the detected energy within a particular filter exceeds a set threshold, then a target is determined to be present with a certain probability. There are several methods to implement the filter bank. With the recent advances in digital signal processing hardware and software, one technique is to structure an algorithm using filter coefficients which are stored within the processors data or program memory.

The determination of range is considerably more complex when relative motion between radar and target is considered. The equation representing the output frequency in this case is: (see appendix A for derivation)

$f_S = 2f_0v/c + 2(\Delta F/\Delta t)R/c + 2(\Delta F/\Delta t)vt/c,$

where,

v is the normal component of relative velocity,
f_0 is the operating frequency, and,
t is time.

The first term is a frequency component contributed by the Doppler effect; the second component is contributed by the range; and the third component is contributed by the change in range, or the velocity component. Note that with no relative motion between radar and target, i.e. v = 0, The equation reduces to $f_S = 2(\Delta F/\Delta t)/c$ as previously presented; if the radar frequency is not changed, i.e. $\Delta F/\Delta t = 0$, only the Doppler component remains, $fs = 2f_0v/c$. This provides a convenient method of measuring the target velocity and possibly some processing options.

2.2 Radar Range Equation

For either a pulsed or FMCW type radar, the maximum range of operation may be determined from a number of system and target parameters when entered into the radar range equation. Due to the statistical nature of the target cross section, weather conditions, minimum detectable signal level, propagation path variations and operational environment, the accuracy of the predicted range when compared with field experiments is generally poor. Notwithstanding the poor accuracy, the radar range equation does provide a basis for parameter comparison and trade studies as well as a model for the system operation.

The power at the radar receiver input is the product of the transmitted power, P_t, the transmitter antenna gain, G_t, the two way path loss, $(1/4\pi R^2)^2$, the effective target cross section, σ, and the effective area or aperture of the receive antenna, A_e. This relationship may be deduced thru the following reasoning:

The power density at a given range from a transmitter = $P_tG_t/4\pi R^2$.

The power reflected from a target = $P_tG_t\sigma/4\pi R^2$.

The power density of the target at the receiver = $P_tG_t\sigma/(4\pi R^2)^2$.

The power intercepted by the receiving antenna, $P_r = P_tG_tA_e\sigma/(4\pi R^2)^2$.

This is the basic form of the radar range equation. The minimum detectable signal at the receiver is designated P_{rmin}, and occurs at the maximum operational range, R_{max}. Substituting into the equation, one may determine the maximum operational range:

$$R_{max} = [P_tG_tA_e\sigma/4\pi P_{rmin}]^{1/4}.$$

There are a number of forms of this equation which make use of the following expression for antenna gain in terms of aperture and operating wavelength, λ, and the antenna physical area, A, and efficiency, η:

$$G = 4\pi A_e/\lambda^2 = 4\pi A\eta/\lambda^2.$$

Equations which are of some interest are repeated here:

$$R_{max} = [P_tA_e^2\sigma/4\pi\lambda^2P_{rmin}]^{1/4}$$

or

$$R_{max} = [P_tG^2\lambda^2\sigma/(4\pi)^3P_{rmin}]^{1/4}.$$

The minimum detectable signal may be approximated as that signal level which engenders a signal-to-noise ratio at the receiver output of 0 dB. The test equipment configuration required to perform this measurement is shown in Figure 2.2.

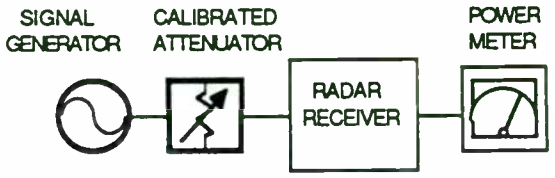


Figure 2.2 Minimum Detectable Signal Measurement

The measurement procedure consists of reducing the input signal level such that the power meter reads only the receiver output noise level. Next, the attenuator is decreased until the power meter reads 3 dB above the noise level. At this input power level, the signal-to-noise ratio is 0 dB and the minimum detectable input power level has been established. The procedure is somewhat more complicated when measuring pulsed radar receivers.

The equivalent input noise level in dBm, which is the same as the minimum detectable signal, may be written mathematically (see appendix B):

$$Ni = KT_oBL_c(F_{if} + T_r - 1),$$

where,

KT_o is Boltzman's constant times absolute temperature,

B is the receiver bandwidth,

L_c is the mixer conversion loss,

F_{if} is the IF amplifier noise figure, and,

T_r is the mixer diode noise temperature ratio.

3.0 SYSTEM IMPLEMENTATION

The system implementation is presented below commencing with a brief discussion of the block diagram and the component development.

3.1 System Block Diagram

The system block diagram is illustrated in Figure 3.1 and depicts the basic elements of the FMCW radar:

- WIP or waveguide integrated package;
- Preamplifier/Filter PCB Assembly;
- Gunn Voltage Controlled Oscillator or VCO;
- Digital Signal Processor PCB Assembly;
- +/- 12 V DC/DC Converters;
- Display Assembly;
- Pyramidal Horn Transmit and Receive Antennas.

The Gunn VCO is employed as both transmit and receiver local oscillator source; a coupler (6 dB) samples the transmit power and applies it to the Schottky diode, quadrature mixer. The receive signal enters the system thru a pyramidal horn and is then applied to the other port of the quadrature mixer. A preamplifier (60 dB gain) boosts the converted signal and applies it to the band pass filter. The amplified/filtered signal then becomes the input to the digital signal processor (DSP) where further filtering and amplification take place. Initially, two algorithms were employed in the digital signal processor: a feedthru mode and a Fast Fourier Transform (FFT) mode. The feedthru mode routes the input signal thru the codec where both filtering and A/D conversion are performed; the signal is then applied to the digital signal processor (ADSP-2101) where the sampled value from the codec is stored/retrieved and subsequently routed back to the codec, processed via D/A conversion and output. In the FFT mode, the codec again supplies the A/D samples to the DSP where the sequence of values (256 values of the sampled input signal) are stored and subsequently multiplied by stored coefficients in accordance with a radix four, complex, FFT algorithm. The output sequence represents the spectral content of the input signal with a resolution bandwidth of approximately 30 Hz (approximately 4 KHz/128).

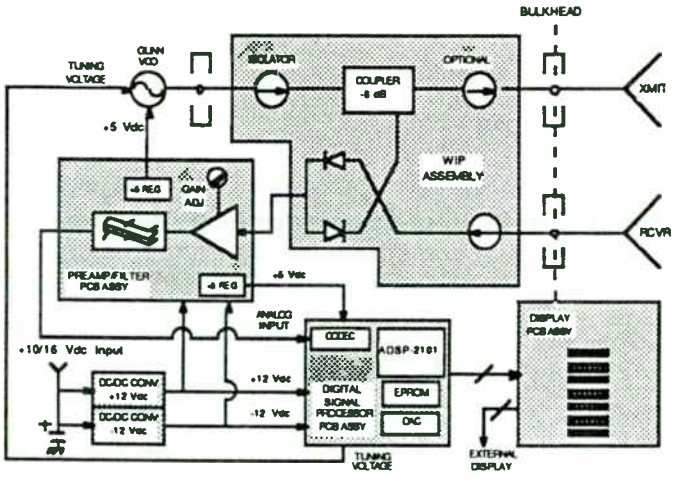


Figure 3.1 System Block Diagram

Each frequency bin of the FFT represents a target range. The values of FFT are sequentially strobed to a D/A converter and oscilloscope. The resulting display is the classic 'A' scope of target amplitude on the vertical axis and target range on the horizontal axis.

Two DC/DC converters have been employed to facilitate operation from a single positive voltage source over a range of input voltages from 10 Vdc to 16 Vdc.

The Unit is housed within an extruded and brazed aluminum enclosure and fitted with a conventional camera attachment for tripod mounting. A low cost LCD portable oscilloscope was utilized for information display purposes. The entire assembly weighs a total of less than 12 pounds.

3.2 Component Development

A number of components were developed for the prototype hardware and for special tests to be conducted on other components and the prototype hardware. In order to provide a comprehensive understanding of the prototype operation, all components of the system will be reviewed and test data will be presented. In some cases, the component parameter data will be discussed relative to the performance of the prototype hardware.

3.2.1 Voltage Controlled Oscillator (VCO)

The VCO is the most critical component of prototype hardware. The VCO supplies both transmitter and local oscillator signals and ultimately, via tuning linearity, determines the radar range resolution performance. A VCO which exhibits a nonlinear tuning characteristic spreads the spectrum of the IF signal. One may immediately discern this through examination of the equation for the IF signal developed earlier. The instantaneous frequency of the IF signal was found to be linearly proportional to the rate of change of frequency with time. Although no evidence was found in the prototype hardware, the VCO PM noise will also contribute to noise in the IF band.

The VCO employed within the prototype hardware was a high quality, waveguide cavity, Gunn oscillator with electronic tuning via hyperabrupt GaAs varactor diode. The equivalent circuit of the VCO is illustrated in Figure 3.2. The Gunn diode is a negative resistance device which is coupled to the waveguide cavity resonator; the varactor diode is similarly coupled. The net loop resistance is negative and thereby becomes a source of energy at the resonator frequency. A positive load resistance is coupled to absorb a portion of the energy. If the load coupling increases to the point where the net loop resistance becomes positive, the oscillation is extinguish.

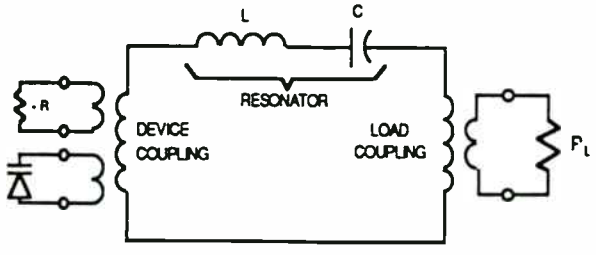


Figure 3.2 VCO Equivalent Circuit

The mechanical configuration of the Gunn oscillator is depicted in Figure 3.3. The waveguide cavity is approximately $\lambda/2$ long at the operating frequency. Both the Gunn and varactor diodes are coupled to the cavity by posts thru the top wall of the waveguide. This configuration has been successfully employed for many years and is the topology of choice for the commercial sources group at M/A-COM. The VCO operates from well regulated +5 Vdc supply at a current of approximately 300 ma.

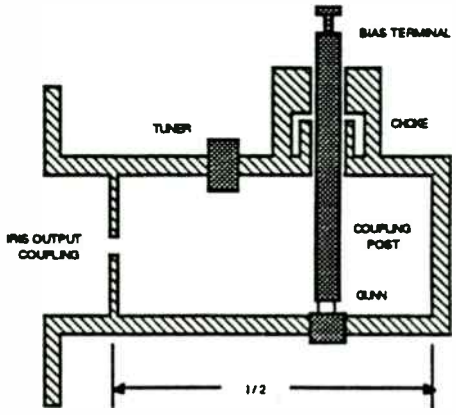


Figure 3.3 Gunn Oscillator Configuration

The power output and voltage tuning characteristic of the VCO is shown in Figure 3.3. Analysis of the data indicates that the deviation from linear tuning is less than one percent over the four to seven volt range.

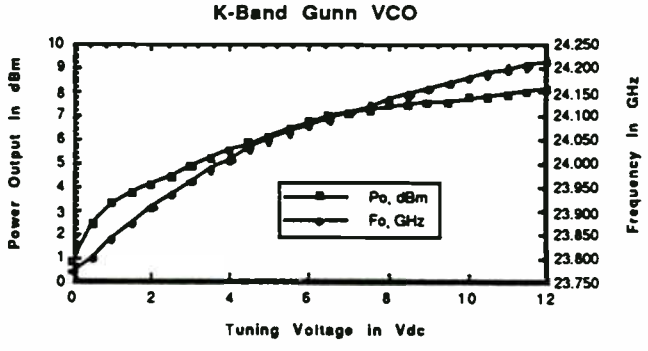


Figure 3.4 VCO Power Output and Voltage Tuning

3.2.2 Waveguide Integrated Package

The Waveguide Integrated Package or 'WIP" is a construction medium used at M/A-COM for several years to achieve optimum component and subsystem performance within a planar physical realization which supplies both the microwave transmission medium and the mechanical housing. The WIP configuration makes use of the extensive numerically controlled machinery and lends itself well to low cost investment or die casting. The WIP construction technique is illustrated in Figure 3.4. The channels are machined or cast and permit the propagation of the TE_{10} rectangular waveguide mode. Semiconductor devices are coupled to the waveguide transmission medium by means of posts;

similar to the Gunn diode mounting discussed in the previous section. The WIP packaging technique allows the integration of a number of components within a single chassis which also serves as a mounting platform or enclosure for various electronic functional controls such as PIN diode drivers, VCO tuning circuits, IF and video amplifiers, voltage regulators, etc.

The WIP package which was developed for the FMCW radar program is illustrated schematically in Figure 3.5. The integrated functions are the LO isolator, the LO coupler, optional transmit and receive port isolators, and a balanced mixer. The transmit and receive port isolators are configured such that a single antenna or separate transmit and receive antennas may be utilized. The balanced mixer employs a 'short slot' quadrature hybrid for the separation of the local oscillator and signal ports. The 6 dB local oscillator coupler is also implemented using a similar narrow waveguide wall coupling aperture. The ferrite isolator originated from a previous design at a slightly lower operating frequency. The isolators are constructed with a teflon-ferrite-teflon 'sandwich' which is placed upon impedance matching triangles at waveguide junctions formed at 120 degrees. The loads for the isolators are simple polyiron material with a machined taper transition in two planes.

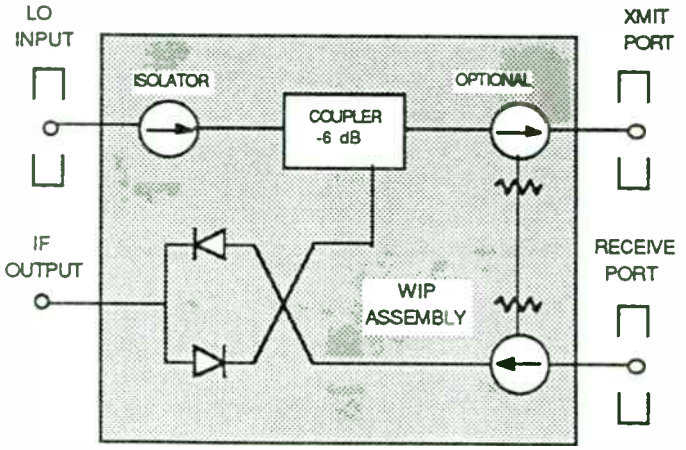


Figure 3.6 WIP Schematic

The performance of the WIP assembly is illustrated in Figures 3.6 and 3.7. Note that the operating bandwidth of the WIP assembly is significantly wider than the required 24.125 +/- 0.10 GHz.

To summarize the performance: the VSWR at all ports is less than 1.20:1, the transmitter to antenna loss is less than 3 dB, and the single sideband noise figure is less than 10 dB over a frequency range from 23.0 to 25.0 GHz. No attempt has been made to improve the noise figure thru diode selection; anticipated improvement is 1.0 to 2.0 dB.

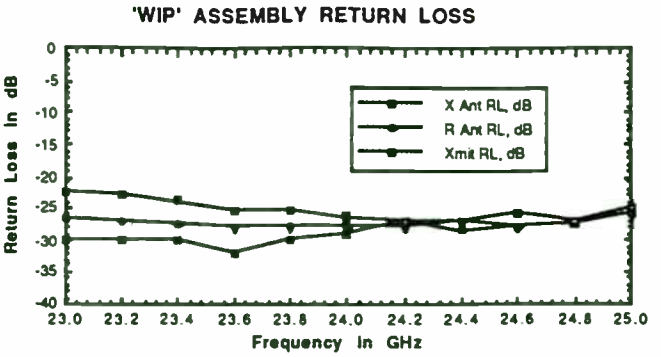


Figure 3.7 WIP Assembly Return Loss

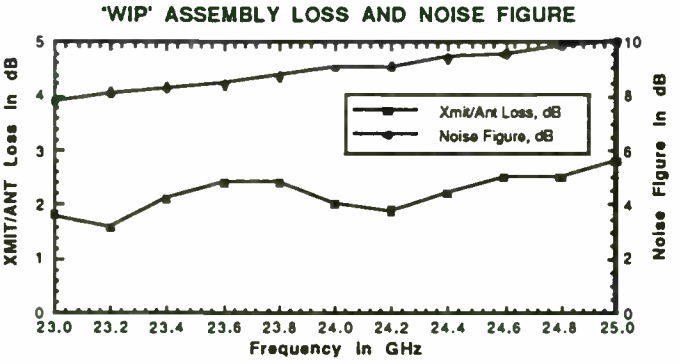


Figure 3.8 WIP Assembly Loss and Noise Figure

3.2.3 Preamplifier/Filter

The preamplifier is required to raise the signal level at the mixer output prior to application to the processor, while the filter is required to limit extraneous signals and noise. Specifically, the filter has a low frequency cut-off of approximately 1.0 KHz to attenuate those signals generated from the frequency sweep. These signals are the result of variations in the mixer output voltage over the sweep frequency. The upper frequency limit of the filter is approximately 5.0 KHz. The CODEC within the digital signal processor also has a fifth order switched capacitor bandpass filter with low and high frequency cut-offs of 200 Hz and 3.4 KHz, respectively. A schematic of the preamplifier filter is illustrated in Figure 3.8.

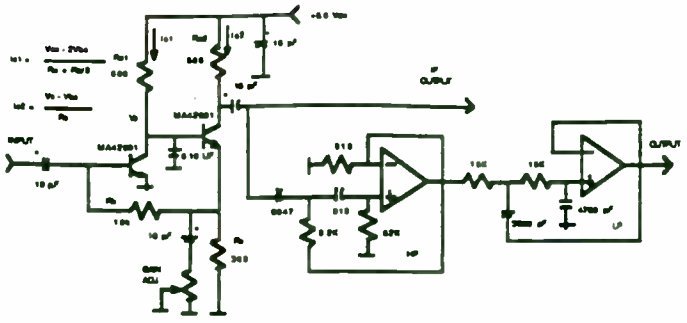


Figure 3.9 Preamplifier/Filter Schematic

The passband response of the amplifier/filter is displayed in Figure 3.9. The gain at 1.0 KHz is 60 dB and the noise figure is estimated at 2.0 dB.

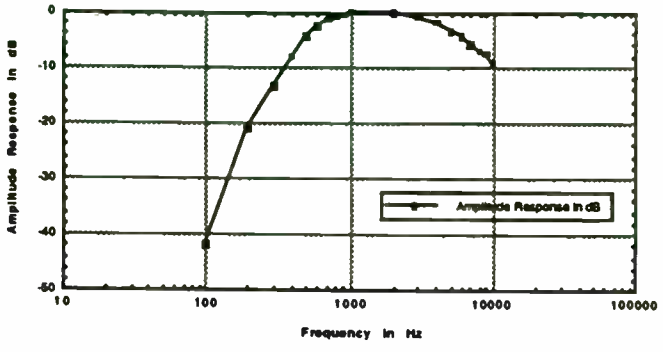


Figure 3.10 Preamplifier/Filter Frequency Response

3.2.4 Digital Signal Processor

To avoid the cost of developing an integrated digital signal processor, it was decided to purchase a general purpose board which incorporates a digital signal processing integrated circuit and the necessary peripheral functions such as sampling, A/D conversion, filtering, D/A conversion, program and data storage, etc. The selected board is the Analog Devices EZ-LAB. The EZ-LAB is a low cost evaluation and demonstration tool for the ADSP-2101 DSP microcomputer. It permits the testing of coded digital signal processing applications on the ADSP-2101. The EZ-LAB is equipped with a 27C512 EPROM containing several prepared processing algorithms. The EPROM is socket mounted and may be replaced with another EPROM containing user created algorithms. The DSP board measures 4.5 X 6.0 inches and is functionally illustrated in Figure 3.10 below.

The DSP board contains a processor controlled CODEC (TP3054) which samples (8 KHz rate) and digitizes (8 bit) the input signal; the resulting data is converted to serial mode and enters the ADSP-2101 digital signal processor. The data is processed in accordance with the algorithm in the program memory of the DSP.

Currently, we have made use of two of the 'canned' algorithms (firmware) supplied as boot pages on the 27512 EPROM. The first algorithm is a feedthru mode whereby the digitized signal enters the DSP serial port from the CODEC. The signal is output from the DSP to the CODEC, back to a D/A converter and fed to the LM388 for amplification and sent to the analog output of the DSP board. It should be mentioned that the CODEC contains a fifth order switched capacitor bandpass filter (200-3400 Hz).

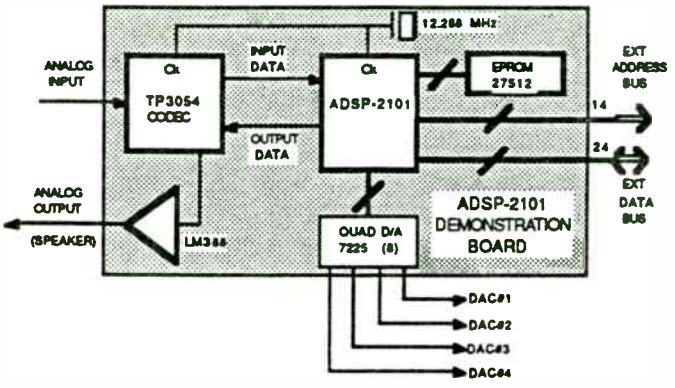


Figure 3.11 Digital Signal Processor Functional Schematic

Another algorithm utilized is a 256 point, radix-4, complex Fast Fourier Transform or FFT algorithm. A frame of data (256 sampled points) is processed and stored in a results buffer; the results buffer is then scanned and displayed on an oscilloscope at 40 frames per second. The timer function of the DSP continuously scans and displays the results and allows updates to the changes in the spectrum. The display driver generates a negative sync pulse to allow proper operation of the oscilloscope. The resulting display resembles a radar 'A' scope type display, i.e., target amplitude and range. The 256 point FFT display presents the frequency or range bins sequentially in a spectrum analyzer or magnitude-squared format.

The DSP board has an operational limitation; the relatively low (8 KHz) sampling rate permits processing signals over a relatively narrow frequency band, i.e., 4 KHz, or half the sample rate. Although this is not a severe restriction for general purpose applications, processing signals with wider bandwidths or Doppler components is not possible without consideration of a more complex algorithm. An alternate DSP board could be obtained with a CODEC operating at a higher sampling rate.

This is one of the recommendations for a possible next phase. Many of the field tests employed this algorithm to obtain data on the FMCW equipment.

3.2.5 Horn Antenna

The transmit and receive antennas were low cost, cast pyramidal horns with E-plane and H-plane aperture dimensions of 1.29 X 1.72 inches, respectively. Using the formulae of Braun (reference 3), the antenna gain was calculated to be 16.5 dB. Subsequent gain measurements of 16.8 db sustained the calculation within acceptable accuracy limits. The antenna beamwidths (H and E plane) were calculated using the formulae of Stimson (reference 4): $\varnothing_h$ = 20 °, and, $\varnothing_e$ = 27 °.

3.2.6 DC/DC Converter

Two DC/DC converters were procured to produce +12 and -12 Vdc for operation of the FMCW equipment. The negative converter permitted field operation of the equipment from an automobile battery via the cigarette lighter plug. Two separate +5 Vdc supplies were developed to operate the Gunn oscillator and the logic devices.

3.2.7 Ramp Voltage Generator

In order to utilize the inherent linearity of the VCO, a swept voltage must be generated which has both a symmetrical and linear up/down sweep. A nonsymmetrical or nonlinear ramp voltage will limit target resolution and accuracy.

A circuit capable of the required linearity and symmetry is illustrated in Figure 3.11. The circuit makes use of a stable oscillator and digital counters with up and down mode control. The counter numeric outputs are cycled through the up and down counts and applied to a D/A converter. The converter output is offset and buffered in order to be compatible with the required VCO control voltage. The circuit further permits ramp frequency control by variation of the modulus of the input counter. The ramp frequency control is required to display full scale ranges of 30, 60, 120, and 240 meters. These ranges are related to ramp frequencies of 100, 50, 25, and 12.5 Hz, respectively.

It is not anticipated that the digital ramp generator be utilized in the final equipment because the DSP is to be programmed to generate the required ramp voltage. However, the linear ramp generator was required for test purposes prior to algorithm development.

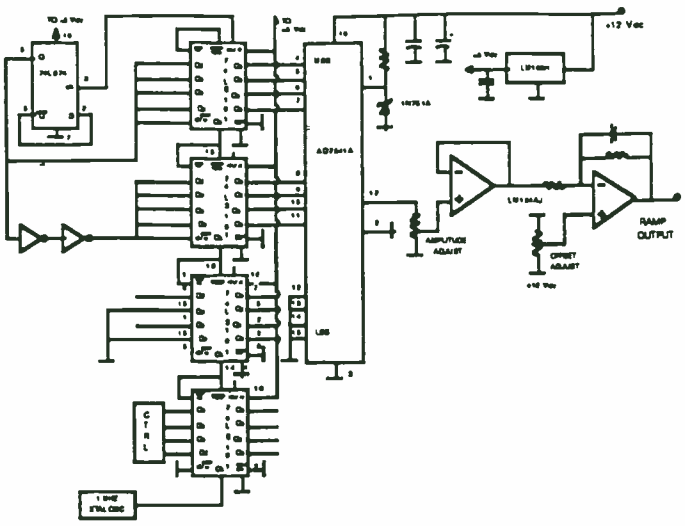


Figure 3.12 Digital Ramp Generator

3.2.8 Display

The display employed during equipment test phase was a purchased portable oscilloscope, Leader Instruments model 100P. The Model 100P is a battery powered combination DMM/oscilloscope with an LCD display. A companion battery powered thermal printer (Model 710) provides hard copies of stored waveform data. The unit measures 5.5 X 4.0 X 1.5 inches and attaches to the FMCW radar equipment back panel.

4.0 SYSTEM TEST

Following assembly and system alignment, both laboratory and field tests were conducted. The test description, conditions and results are presented in this section.

4.1 Laboratory Test

Two configurations of test equipment were utilized to determine the operational quality of the engineering model. The configurations and test results are presented.

4.1.1 Sensitivity measurements

The sensitivity measurement test employed a double sideband generator to offset the transmitter output of the FMCW module; the transmitter frequency was not swept during these tests. An audio oscillator was used to vary the sideband frequency to permit testing of minimum detectable signal, equivalent input noise and IF bandwidth.

The test equipment configuration of Figure 4.1 was utilized to conduct the tests. The transmit frequency was held constant for these measurements.

Using calibrated attenuators and having previously measured the transmitter output power and the sideband conversion loss, one may reduce the power at the input to the receiver and monitor the IF power with the RMS voltmeter and thereby determine the signal to noise ratio within the IF bandwidth for any input power.

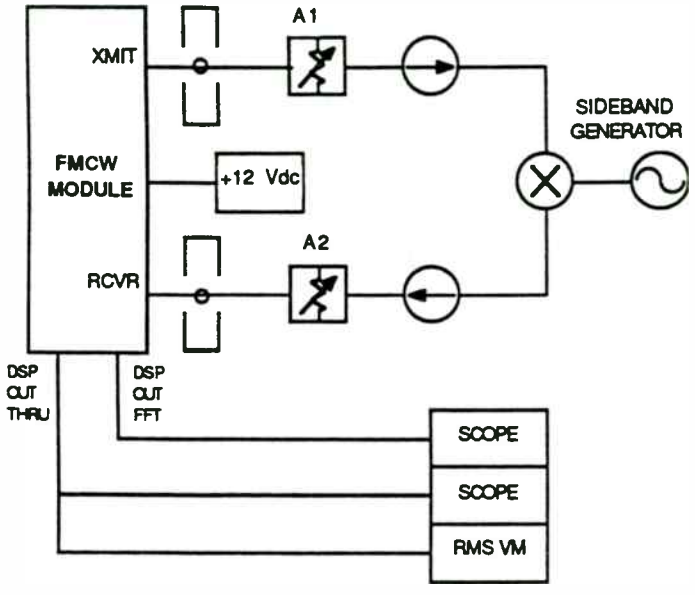


Figure 4.1 Sensitivity Measurement Test Equipment

The IF bandwidth was measured by varying the offset generator frequency while observing the RMS voltmeter and noting the -3 dB points. The following data was recorded:

Transmitter Power	+8 dBm
Sideband Conversion Loss	-10 dB
Attenuator A1	-50 dB
Attenuator A2	-70 dB
Received Power for S/N=0 dB (IF BW = 3 KHz)	-122 dBm

Recalling the mathematical expression for equivalent input noise (appendix B):

$$N_{ie} = L(F_i + T_r - 1)kTB = E\text{-}12.2$$

The diode noise ratio may be calculated; Tr=7.39, or 8.5 dB excess noise.

4.1.2 Operational Test

Operational testing of the engineering model was conducted using the test equipment configuration depicted in Figure 4.2.

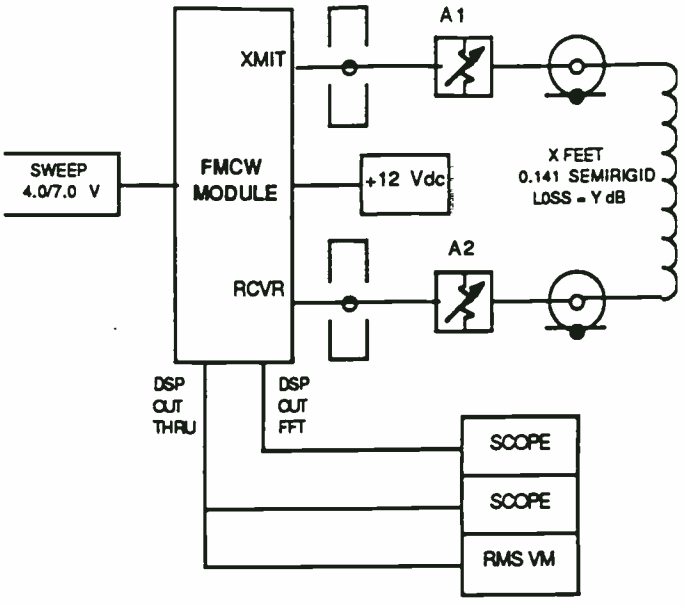


Figure 4.2 Operational Test Equipment Configuration

The operational mode testing simulated targets at various ranges and cross sections by inserting different path attenuations and lengths of 0.141 inch diameter semirigid cable between the transmitter and receiver ports. Documentation consisted of through mode and FFT mode photographs of the IF signal and 256 point FFT spectral output. The FFT spectral output was via frequency (range) bin sequential strobing.

The assorted conditions are set forth in Table 4.1. The oscillograph data is displayed in Figure 4.3.

Table 4.1 Operational Test Conditions

CONDITION NO	PH CABLE LENGTH M	EL CABLE LENGTH M	RANGE SCALE M	SWEEP RATE HZ	XMIT ATTN dB	RECEIVE ATTN dB	RECEIVE POWER dBm
1	3.05	4.42	30	100	35.0	35.4	-70.4
2	9.14	13.25	30	100	35.0	28.4	-79.4
3	9.14	13.25	60	50	35.0	28.4	-79.4
4	18.29	26.50	30	100	11.7	35.4	-85.1
5	18.29	26.50	60	50	11.7	35.4	-85.1
6	18.29	26.50	120	25	6.1	28.4	-72.5
7	18.29	26.50	240	12.5	0.4	25.5	-63.9
8	27.43	39.75	30	100	8.7	23.0	-93.7
9	27.43	39.75	60	50	8.7	23.0	-93.7
10	27.43	39.75	120	25	1.0	23.0	-66.0

4.2 Field Test

Field testing of the engineering model was conducted using the pyramidal horn antennae and calibrated attenuators at both transmit and receive ports. Two triangular trihedrals of known target cross section were employed. A diagram of the triangular trihedrals is shown in Figure 4.3. The triangular trihedral is quite popular as a radar target cross

section because the azimuth and elevation lobes are wide, i.e. constant cross section over angular deviations, and therefore exhibit large skew tolerance.

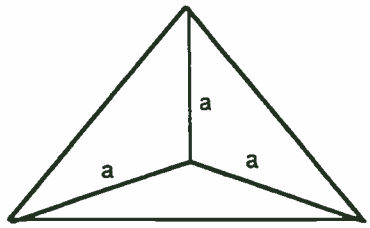


Figure 4.4 Triangular Trihedral

The radar cross section is calculated from the formula (reference 5):

$$\sigma = 4\pi a^4/3\lambda^2.$$

The target cross sections are 3.344 and 33.078 square meters.

The photographic documentation of the field test are displayed in Figure 4.4.

5.0 SUMMARY AND CONCLUSIONS

The program, as originally proposed, was conducted to determine feasibility of the reference 1 system architecture to function as a a general purpose ranging. The program objectives were altered to include the manufacture of an engineering prototype FMCW radar system with more general operational capability.

The engineering model exhibited excellent operational characteristics as evidenced by the data reported herein and demonstrated the requirement for further study and feature enhancement in the following areas:

- system display
- advanced DSP algorithm
- receiver automatic gain control

One conclusion that may be declared without reservation is that the FMCW radar concept is appropriate for a wide range of potential applications and that the RF architecture is simple and cost effective; the addition of digital processing techniques and algorithm development will likely be the vehicle which generates broad market acceptance to the product.

6.0 APPENDICES

Appendix I FMCW IF Signal Development

Appendix II Noise Figure Development

7.0 REFERENCES

1. M/A Technical Memorandum of 10 February 1991
2. Skolnik, M.I.: "Radar Handbook,"McGraw-Hill Book Co, New York, 1970
3. Braun, E.H., Data for the Design of EM Horns, IRE Transactions on A & P, January, 1956.
4. Stimson, G.W., "Introduction to Airborne Radar," Hughes Aircraft Co, El Segundo, CA, 1983.
5. Currie, N.C. and Brown, C.E.,"Principles and Applications of Millimeter-Wave Radar," Artech House, Norwood, MA, 1987.

APPENDIX A
IF SIGNAL DEVELOPMENT

Consider the equipment configuration of Figure A.1 where a VCO is swept at the rate of K Hz/sec by the applied sawtooth signal. The VCO signal is radiated toward a target X at range R and normal component of velocity, v. A portion of the VCO signal is coupled to a mixer LO port; the mixer receives the reflected VCO signal at the RF port. The difference frequency is available at the IF port of the mixer.

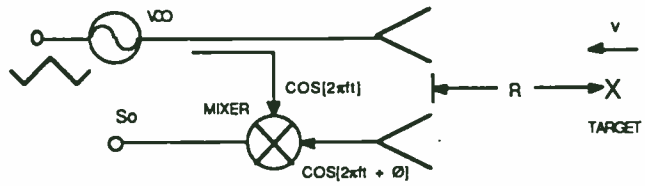


Figure A.1 FMCW Radar IF Signal

The signal, So, at the mixer IF port is the filtered product of the cosinusoidal signals entering the LO and RF ports of the mixer, and may be written:

$$So = A\,COS[\emptyset],$$

where, ø is he total phase shift associated with the two way range, R, i.e., $\emptyset=2[2\pi R/\lambda]$, and A is the attenuation associated with the two way path loss and the target cross section. λ is the transmission wavelength, $\lambda=c/f$. Substituting, one may write:

$$\emptyset = 4\pi Rf/c.$$

Now remember that the frequency is linearly changing at the rate K, and that the sawtooth starts at frequency, fo, and therefore ø may be written:

$$\emptyset = [4\pi R/c]*[f_O + Kt], \text{ or,}$$

$$\emptyset = 4\pi Rf_O/c + 4\pi RKt/c.$$

From FM theory, the instantaneous frequency is defined as:

$$f_i = [1/2\pi]*[\partial\emptyset/\partial t].$$

Performing the differentiation:

$$f_i = [2f_o/c]*[\partial R/\partial t] + [2K/c]*[R + t\partial R/\partial t].$$

Rewriting the equation with the substitution, $\partial R/\partial t = v$:

$$f_i = 2f_o v/c + 2KR/c + 2Kvt/c.$$

Note that there are three discrete terms which describe the instantaneous frequency at the mixer output:

1. $2f_o v/c$, the 'Doppler' frequency component;

2. $2KR/c$, the frequency component associated with target range, R;

3. $2Kvt/c$, the frequency component associated with the change in range.

The following points should be noted:

If there is no relative motion between the radar and target, i.e. $v = 0$, then $f_i = 2KR/c$, f_i is range dependent only.

and,

if the radar frequency is not changing, i.e. $K = 0$, $f_i = 2f_o v/c$, the 'Doppler' component and f_i is velocity dependent only.

APPENDIX B
NOISE FIGURE DEVELOPMENT

Consider the equipment block diagram of Figure B.1. A mixer with conversion loss, L_c, and diode noise temperature ratio, T_m, is connected to an amplifier of gain, G, noise figure, F_i, and noise temperature, T_i. The mixer is driven by an ideal local oscillator, i.e., one that contributes no additive noise to the system. A power meter measures the total output noise in a bandwidth, B, under the input conditions of connection to a termination at temperature, T_a, or a termination at temperature, T_b.

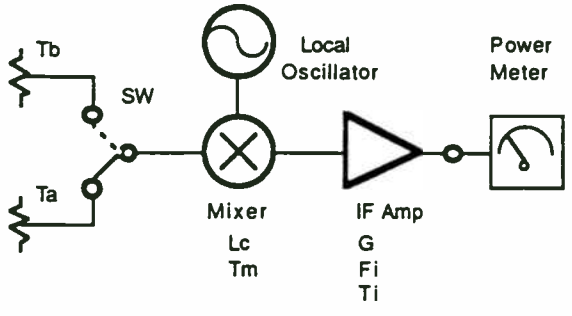


Figure B.1 Noise Figure Measurement

The Y-factor is defined as the ratio of the total noise power output under the two switch conditions, i.e., Y=Nb/Na. Now, one may write,

$$Na = F_i kT_a BG + k(T_m - T_a)BG, \text{ and,}$$

$$Nb = Lk(T_b - T_a)BG + FikT_a BG + k(T_m - T_a)BG,$$

where, k is Boltzman's constant, 1.38exp(-23) Joules/°K.

Forming the Y-factor and eliminating the common multiplier, kBG, one obtains:

$$Y = [L(T_b - T_a) + F_i T_a + (T_m - T_a)]/[F_i T_a + (T_m - T_a)].$$

Subtracting -1 from both sides of the equation, one obtains:

$$Y - 1 = [L(T_b - T_a) + FT_a + (T_m - T_a)]/[F_i T_a + (T_m - T_a)] - 1.$$

Performing some algebraic manipulation:

$$Y - 1 = L(T_b - T_a)/[F_i T_a + (T_m - T_a)].$$

Remember that Fs, system noise figure, has the following mathematical relationship:

$$Fs = (T_b - T_a)/(Y-1)T_a, \text{ and } (T_b - T_a)/T_a = ENR.$$

Substituting into the equation for Fs, one obtains:

$$Fs = (F_i + T_m/T_a - 1)/L.$$

Tm/Ta is called the diode noise temperature ratio, Tr, and upon substitution and exchange of the conversion loss, L, from a numeric which is less than one to an inverse which is greater than one, the formula may be written in the more conventional manner:

$$F_s = L(F_i + T_r - 1)$$

The noise figure represents the degradation in signal to noise ratio as a signal progresses thru a system; the formal definition is:

$$F_s = [S_i/N_i]/[S_o/N_o].$$

Rearranging terms and with the following substitutions: Ni=kTB and So/Si = G, yields:

$$F_s = N_o/kTBG.$$

The equation leads to another(equivalent) noise figure definition:

> Noise figure is the ratio of total output noise power to that output noise power engendered by the input.

The total output noise power, N_o, may be written in terms of the system noise figure:

$$N_O = F_S kTBG.$$

The equivalent input noise is:

$$N_{ie} = F_S kTB.$$

Substituting the expression developed for the system noise figure, F_S, the equivalent input noise, or minimum detectable signal, is:

$$N_{ie} = L(F_I + T_r - 1)kTB.$$

Designing RF Circuits and Modules using Modern CAD Tools

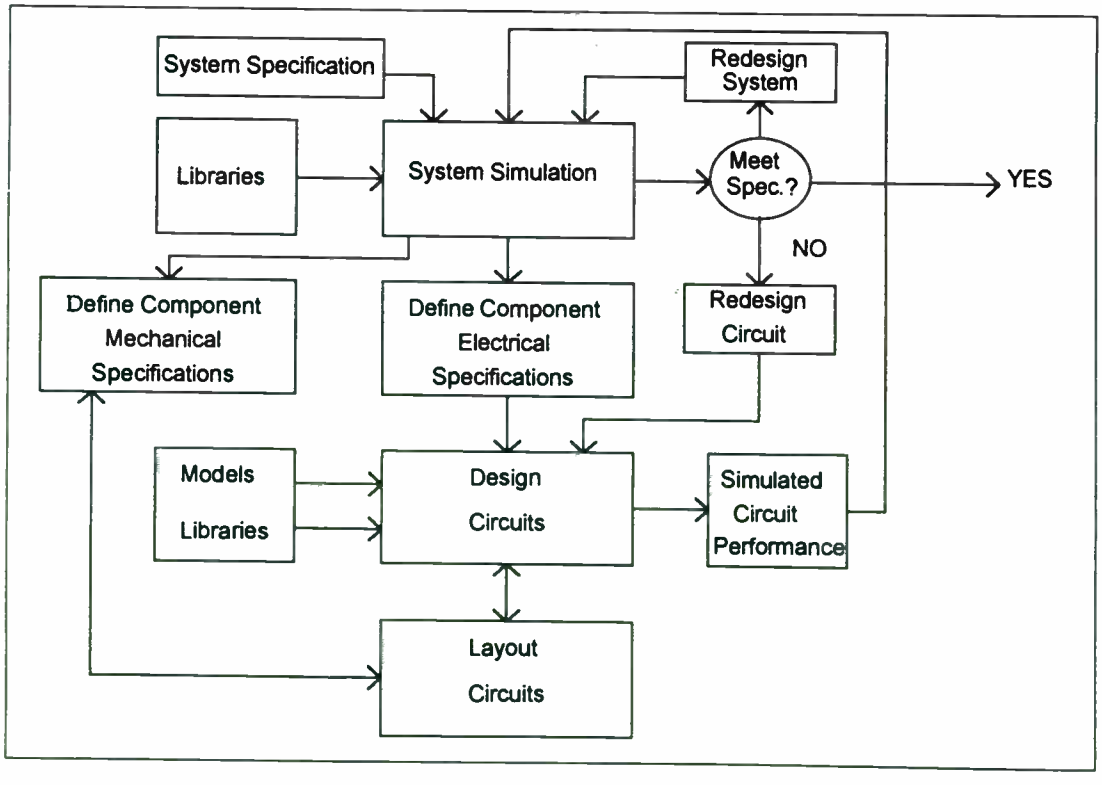
Raymond S. Pengelly
Compact Software, Inc.
Paterson, New Jersey,
USA

Session E-3: Practical RF CAD
RF Expo East, Tampa, FL
October 21

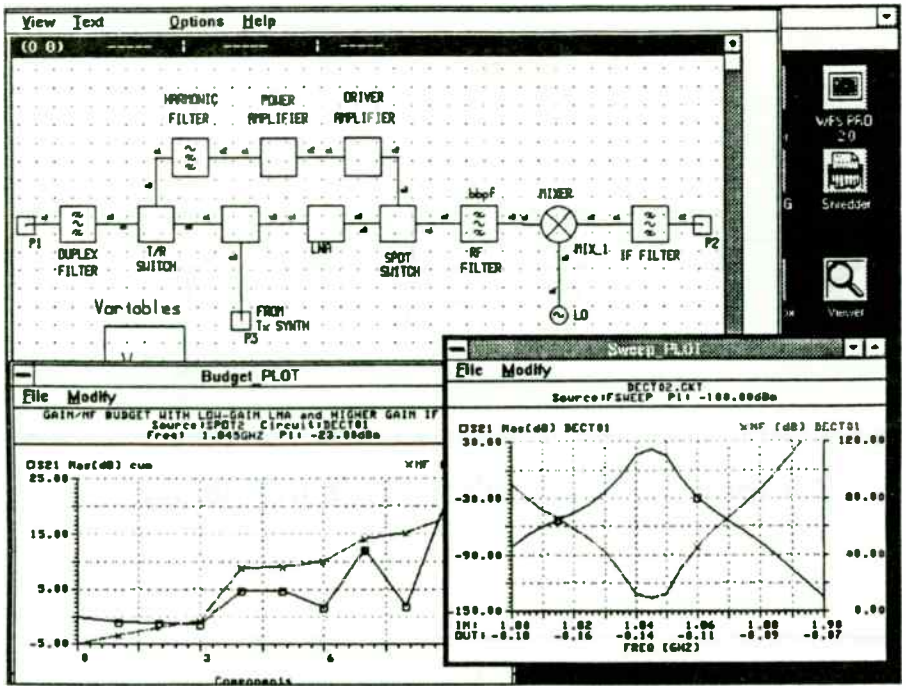
SUMMARY

- ◆ CAD tools for the RF community have reached a high level of sophistication during the last few years
- ● Design complex sub-systems using a software suite, like Compact Software's Serenade®, without leaving the CAD tool environment.
  - ⇨ Modular combination of schematic entry, simulation engine and (semi)-automatic layout preparation available in either PC or workstation-based formats.
- ● Use a System Level Simulator, like Microwave Success®, with a circuit-level linear/nonlinear simulator, like Microwave Harmonica®
  - ⇨ Inputs to the simulators are defined using the Serenade Schematic Editor and the planar integrated/hybrid circuits are defined using the Layout Editor.
  - ⇨ Schematic entry, via Serenade, provides a fast, comprehensive and error-free method of entering circuit and system-level information to the simulators and the layout editor.

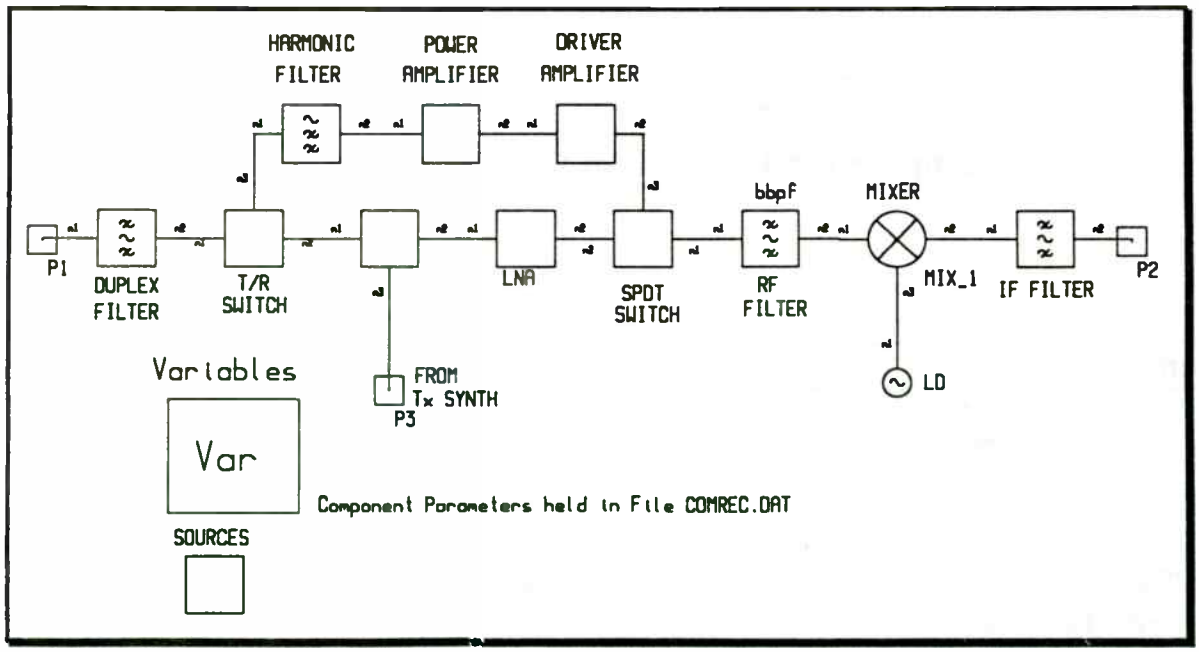
- **System simulation allows the design engineer to investigate a complete range of parameters including:**
  - ✓ **Gain**
  - ✓ **Noise performance**
  - ✓ **Group delay**
  - ✓ **Power compression characteristics**
  - ✓ **Intermodulation distortion**
  - ✓ **Degradation to modulated carriers such as 64-QAM**
  - ✓ **Bit-Error Rate**
  - ✓ **Eye-Diagrams**
  - ✓ **Constellation Plots**
  - ✓ **Phase Noise**



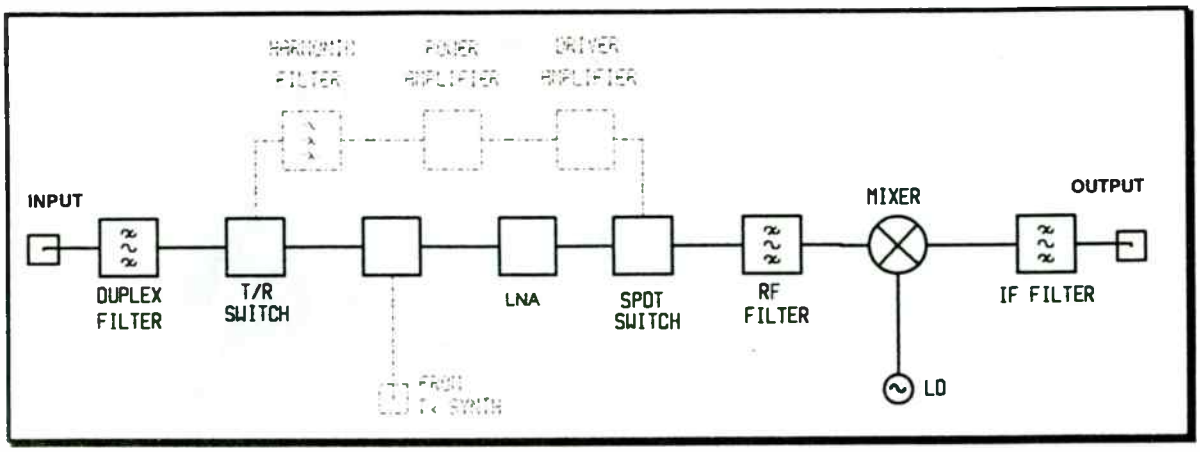
Designing Modules and Systems using Modern CAD Tools



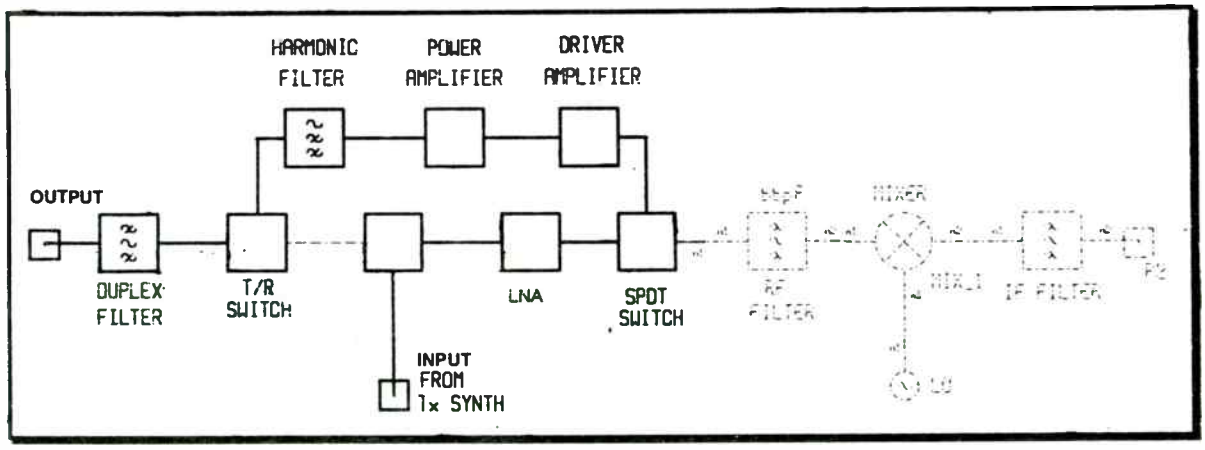
■ **Serenade Schematic Editor and Microwave Success System Simulator running concurrently under Windows 3.1 on PC**



Typical System Diagram for the Transmit/Receive Section of a Digital European Cordless Telephone -- DECT 1800



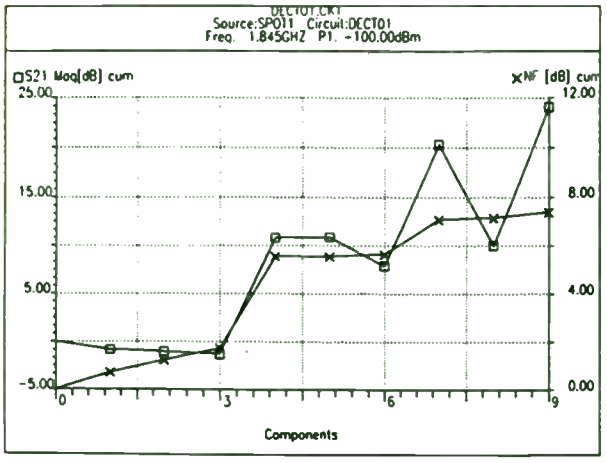
Components Used when CT is in Receive



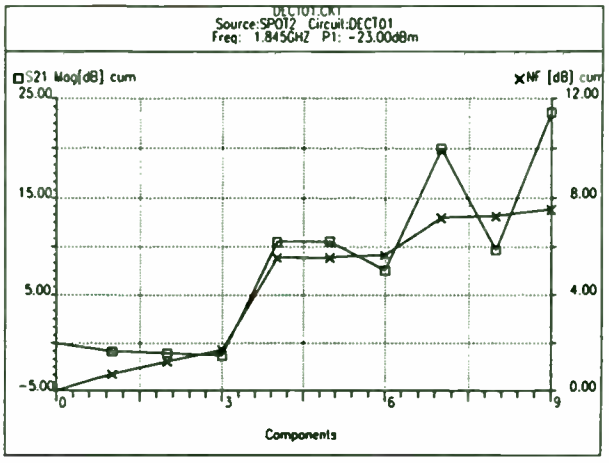
Components Used when CT is in Transmit

Parameter	Specification
Dynamic Range at the antenna	-100 to -23 dBm
Signal-to-Noise Ratio at the Output	min. 9 dB
Noise Figure	max. 12 dB
Third Order Intercept Point at the input	min. -20.5 dBm
In-band Blocking	min. 80 dB
Out-of-band Blocking	min. 106 dB

Specifications for DECT 1800 Front End



(a)

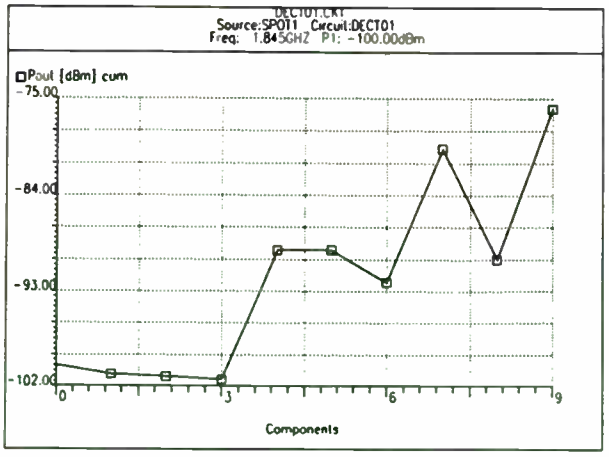


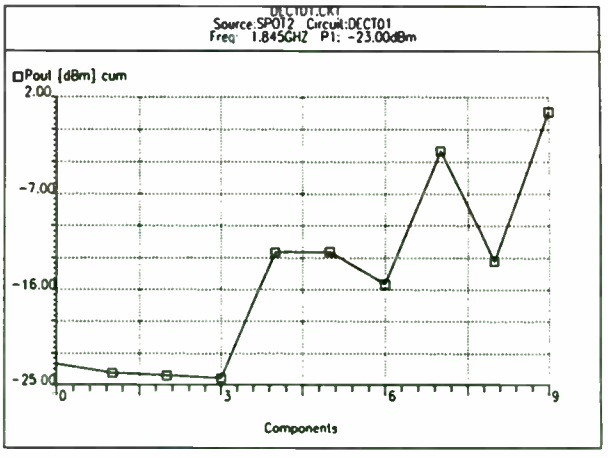
(b)

- **The Microwave Success System Simulator is used to investigate design specification trade-offs.**

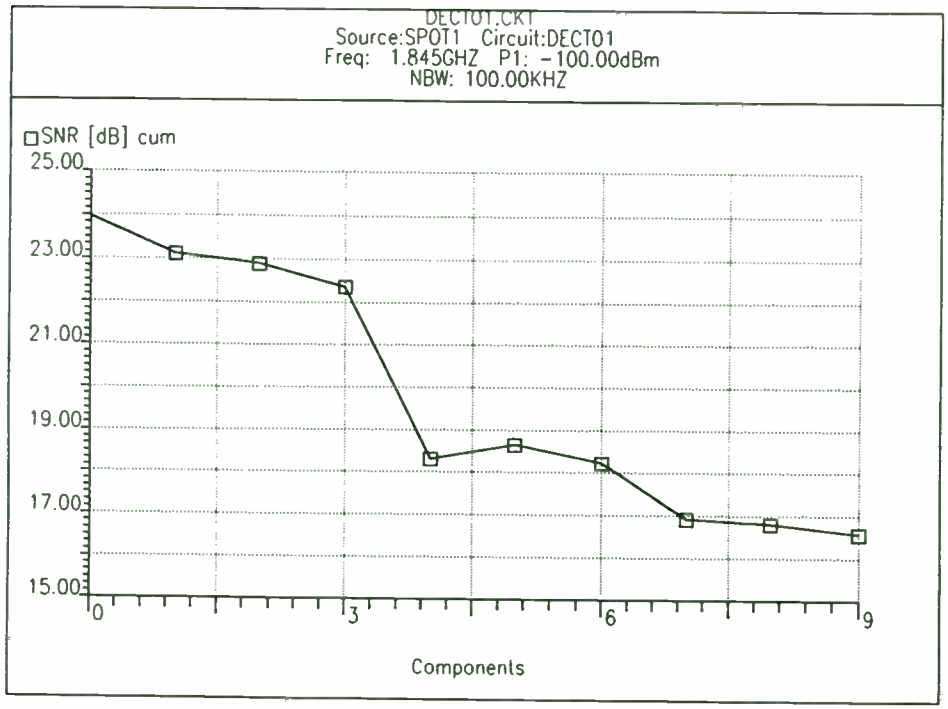
- (a) shows, for example, the results of calculating the system budget at a received signal level of -100 dBm at a frequency of 1845 MHz. The cumulative budget plot shows the way in which the overall gain and noise figure of the front-end "build-up" through the various components. The overall gain of the front-end in this case is 22 dB with a SSB noise figure of 7.6 dB.

- The same front-end, operating with a received signal level of -23 dBm (the specification maximum), has an overall gain of 21.7 dB with a noise figure of 7.7 dB (Figure (b)).

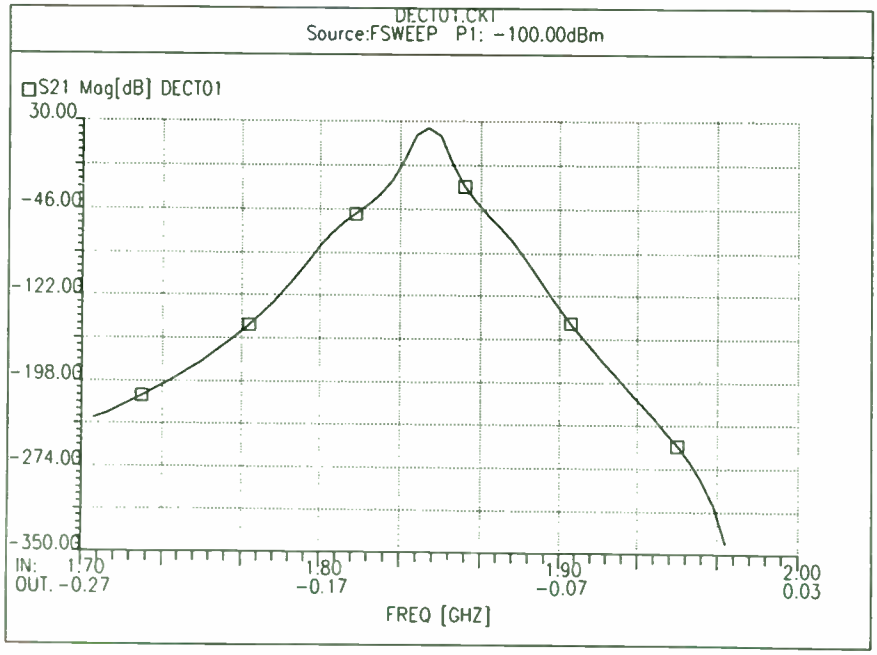




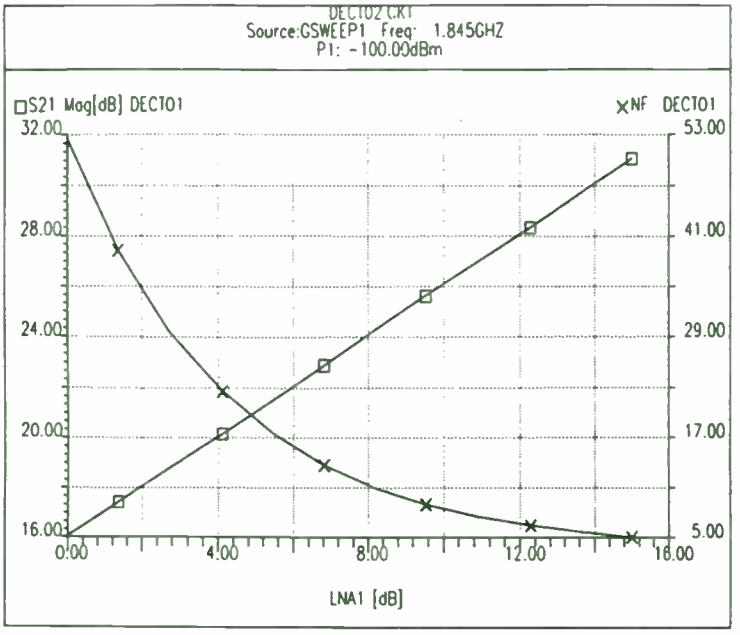
- **Microwave Success calculates the overall performance of the system and displays the results using a variety of graphic displays. Shown above is the output power of a DECT mobile station on receive displayed as a cumulative budget plot at two different receive levels, -100 dBm and -23 dBm.**



- **The overall signal-to-noise ratio (SNR) for DECT 1800 must be greater than 9 dB. With an input signal of -100 dBm the SNR is 16.2 dB**

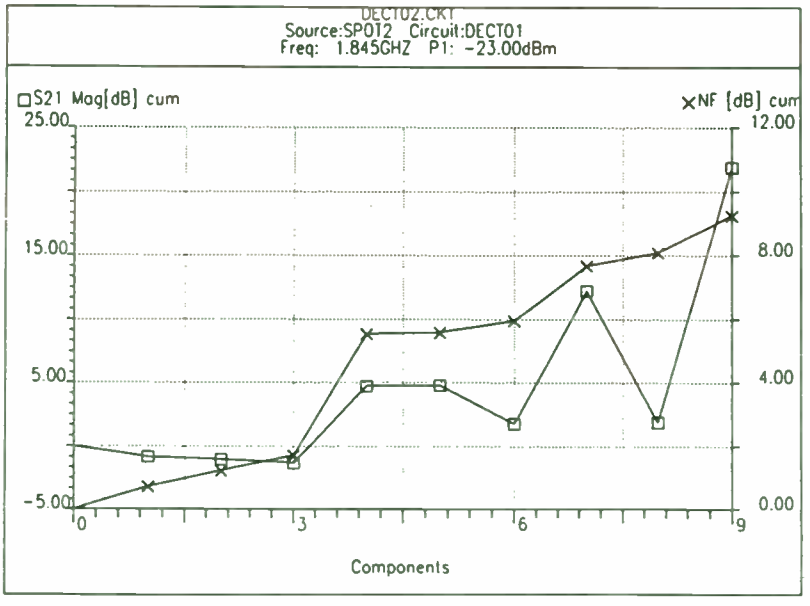


- **The overall response of the front-end at a fixed local oscillator frequency of 1975 MHz is shown. The plot clearly shows the effects of the various band-pass filters in the front-end which aid in meeting the in-band blocking and out-of-band blocking specifications.**



Methods of achieving the required system specification that result in rather different (and, perhaps, unrealistic) specifications for the single-pole, double-throw switch, LNA, filters and mixer can be investigated. For low-cost production it is mandatory that the performance of each circuit is reproducible and easy to achieve resulting in high yield.

■ DECT 1800 SystemTrade-Off Analysis using Microwave Success



In order to achieve a front-end noise figure of 12 dB, an amplifier with a gain of 6 dB is needed (mixer noise figure and gain of 10 dB). Increasing the gain of the amplifier to 14 dB gives the required overall front-end gain but may not be the most cost-effective method.

■ **Above we show the case where the front-end gain is mainly concentrated after the mixer in the first IF amplifier -- in both cases the noise figure of the LNA was 4 dB. In this case the noise figure of the mixer was made 6 dB, which is easily achievable with active mixers.**

◆ **DECT 1800 System Trade-Off Analysis using Microwave Success**

Component	Gain	NF	IP3 (input)	Power
Duplex Bandpass Filter	-3.0 dB max.	<3.0 dB	150 dBm	---
SPDT T/R Switch	-0.5 dB	0.5 dB	>0 dBm	---
SPDT T/R Switch	-0.5 dB	0.5 dB	>0 dBm	---
Low-Noise Amplifier	14.0 dB	<6.0 dB	>0 dBm	---
SPDT Switch	-0.5 dB	0.5 dB	> 0 dBm	---
RF Filter	-4.0 dB max	<4.0 dB	150 dBm	---
Mixer	10.0 dB	10.0 dB	> 0 dBm	---
Local Oscillator	---	---	---	-10 dBm
IF Filter	-10.0 dB max	<10.0 dB	150 dBm	---

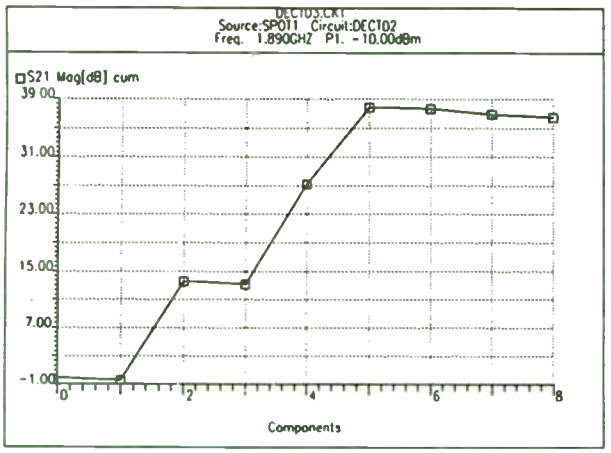
List of the important parameters for the components in a DECT front-end (for Receive)

Parameter/ Component	Value	Gain	Noise Figure	1 dB O/P Power	Third Order Intercept Point @ Input
Frequency range	1880MHz to 1900MHz	---	---	---	---
Power Amplifier	---	37 dB min.	---	26 dBm	---
Low-Noise Amplifier	---	12 dB	4 dB	---	-10 dBm
T/R SPDT Switch	---	-0.5 dB	---	>27 dBm	---
Power from TX Synth.	-10 dBm	---	---	---	---
Power Supply	3.3 to 4.5 volts, three NiCd batteries	---	---	---	---
Temperature range	-10C to +50C	---	---	---	---

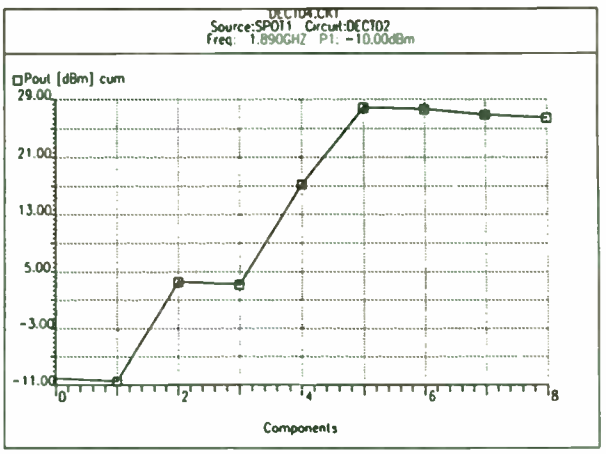
List of the important parameters for the Components in a DECT Front-End (for the Transmitter)

Component	Gain	Noise Figure	Return Losses	P 1 dB comp.	O/P IP3
Duplex Filter	-0.5 dB	0.5 dB	>15 dB	>150 dB	>150 dBm
SPDT Switches 1 & 2	-0.5 dB	0.5 dB	>15 dB	>10 dBm	>20 dBm
LNA	14 dB	4 dB	>20 dB	11 dBm	23 dBm
Driver Amplifier	15 dB	4 dB	>20 dB	16 dBm	28 dBm
Power Amplifier	12 dB	4 dB	>15 dB	>25 dBm	37 dBm
T/R Switch	-0.5 dB	0.5 dB	>15 dB	>30 dBm	>37 dBm
Harmonic Filter	-0.5 dB in band -20 dB at 1900 MHz	0.5 dB	>15 dB	>150 dBm	>150 dBm

Parameters of Components used in system level simulation of DECT Transmitter

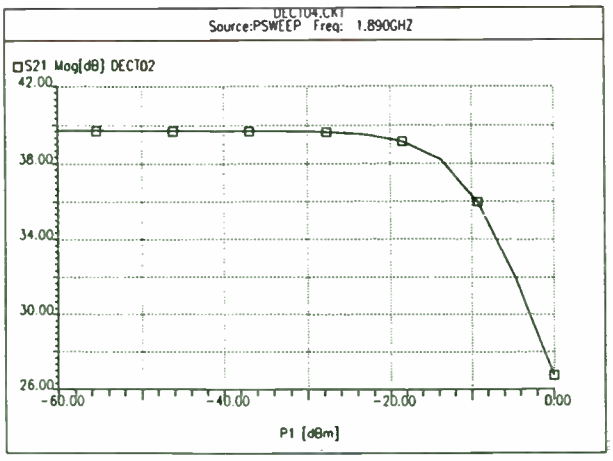


(a)

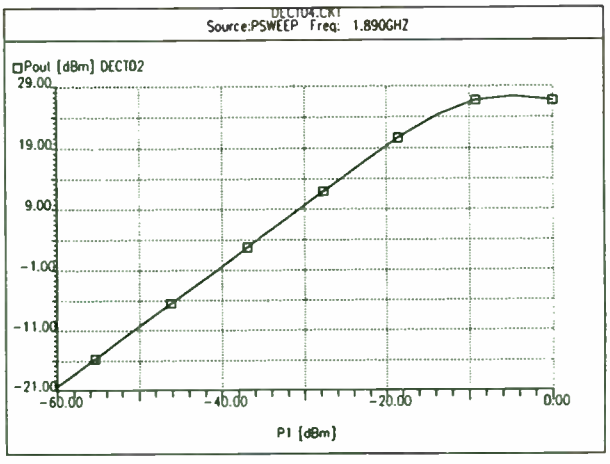


(b)

- (a) shows a budget plot of the gain of the transmit section when the input power level is set at -10 dBm. The overall gain is close to 40 dB.
- (b) shows the corresponding power cumulative budget plot for the transmitter. The power output from the power amplifier is 26.7 dBm. At the antenna the power output level is 25.4 dBm.



(a)



(b)

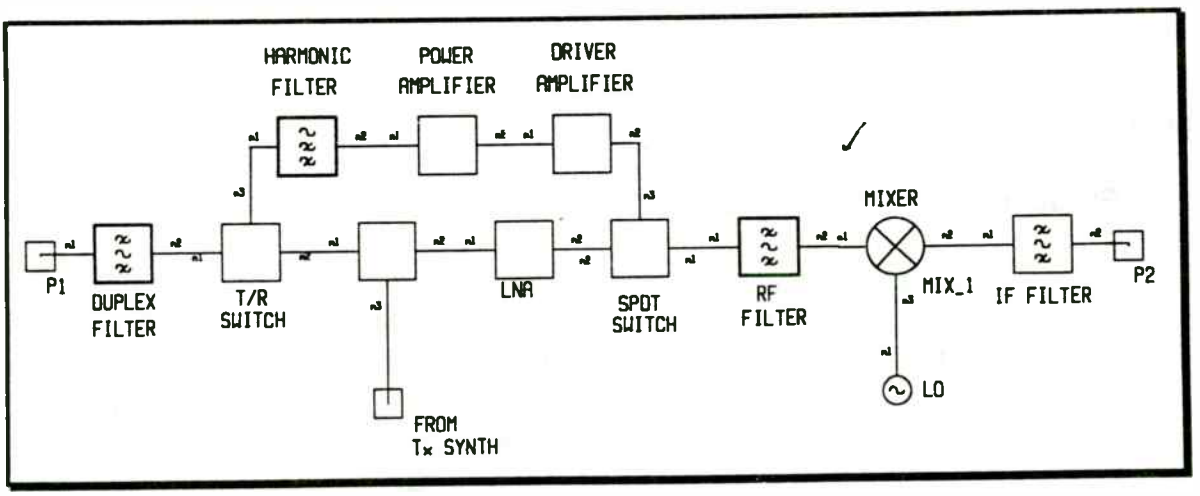
- **The compression characteristics of the transmitter are shown in (a) and (b).**
- In (a) the Input Power to port P3 is swept from -60 dBm to -10 dBm. The Output Power at -10 dBm input is 26.5 dBm with an overall Gain Compression of 3 dB.
- (b) shows the Power Compression curve for the Transmitter.

Linear and Nonlinear Circuit Design

- **Super-Compact provides state-of-the-art linear circuit simulation facilities**
  - *Fast Nodal Analysis on both Workstations and PCs*
  - *Unrestricted Nodal Noise Analysis*
  - *Very Accurate Active and Passive Component Models*
  - *Quasi-Full Wave Analysis Module for Multiple-Coupled Lines*
  - *Complete Libraries of commonly-used components including*
    Bipolar and Field-Effect Transistors
  - *"Electronic" Smith Chart for easy impedance matching*
  - *Circuit Optimization and Tuning*
  - *Statistical Yield Analysis and Circuit Design Centering*
  - *Voltage "Probing" of the internal nodes of circuits*

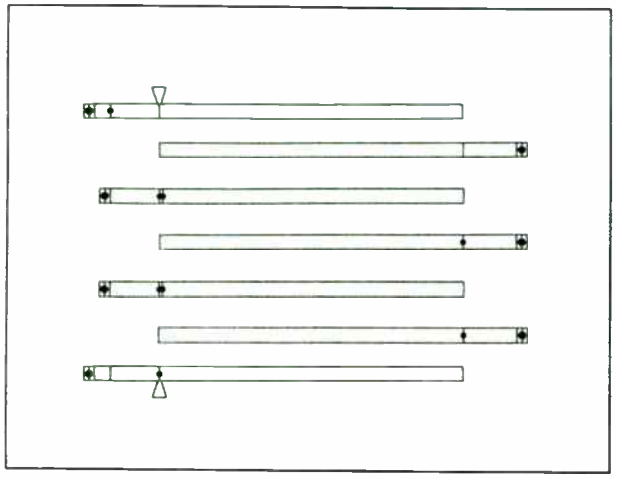
Linear and Nonlinear Circuit Design

- **Microwave Harmonica provides state-of-the-art nonlinear circuit simulation facilities**
  - *Fast Harmonic Balance Technique on both Workstations and PCs -- many times faster than Spice.*
  - *Very Accurate Active and Passive Component Models*
  - *Superior Large-Signal Models such as modified Gummel-Poon for Bipolar Transistors; modified Materka and Triquint-Own-Model (TOM) for FETs and Charge-Conservation Model for MOSFETs*
  - *Complete Libraries of commonly-used components including Bipolar and Field-Effect Transistors*
  - *Circuit Optimization and Tuning*
  - *Statistical Yield Analysis and Circuit Design Centering*
  - *Voltage "Probing" of the internal nodes of circuits*
  - *Many display types including spectral and time domain, power, harmonics, DC voltages and currents*

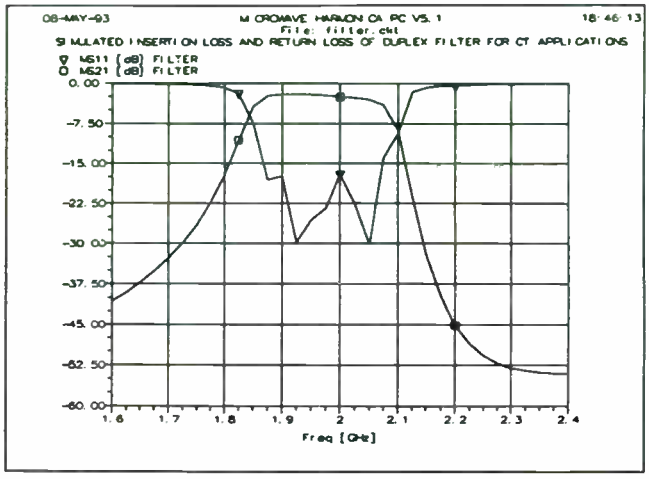


Designing Filters for the CT

- In order to reduce size and keep manufacturing costs low design engineers are adopting novel techniques to realize circuits and sub-systems.
- As RF circuits are packed closer together parasitic, unwanted coupling and cross-talk can take place between components.
- Filters are used to define signal bandwidths, produce image rejection as well as prevent transmission signal harmonics from reaching antennas. These filters need to be physically small and often need high rejection close to the operating signal band.



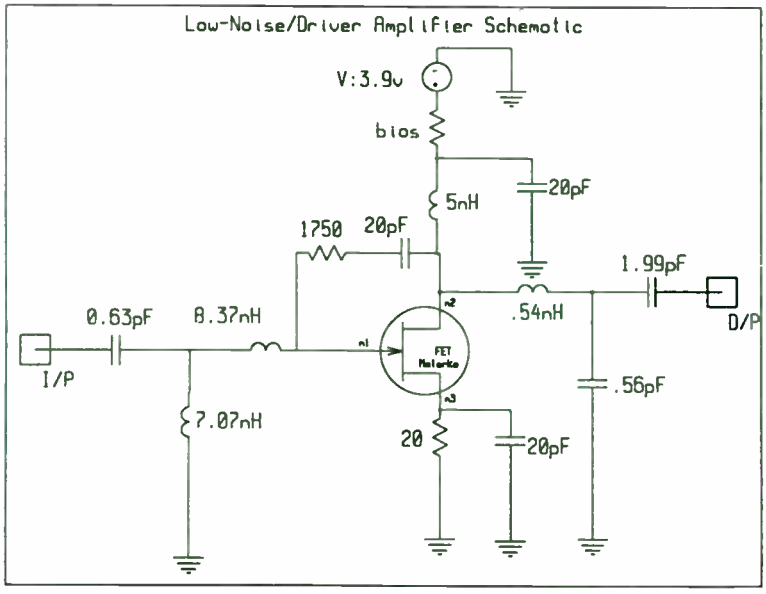
Layout of a Typical Interdigitated Filter



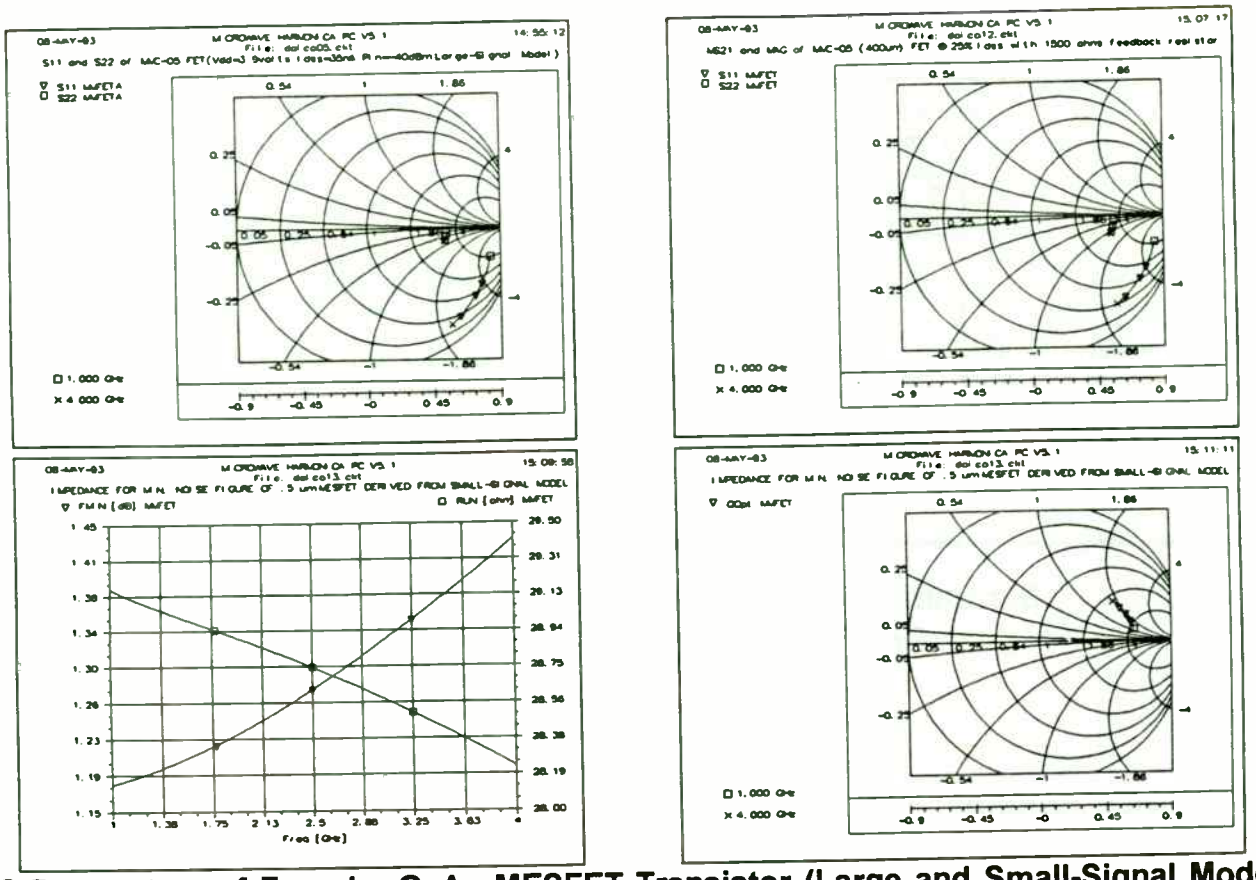
The Simulated response of such a Filter in the 1.8 to 2.0 GHz frequency range.

Linear and Nonlinear Circuit Design

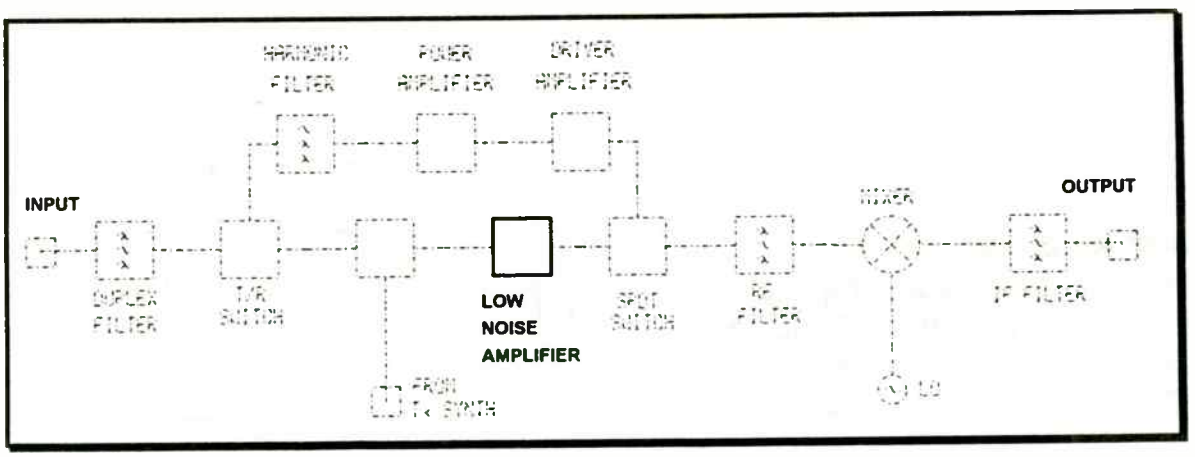
- **Use of Frequency Domain Circuit Simulators, like Super-Compact, for Small-Signal, Linear Design**
- **Use of Frequency/Time Domain Circuit Simulators, like Microwave Harmonica, for Large-Signal, Nonlinear Design**



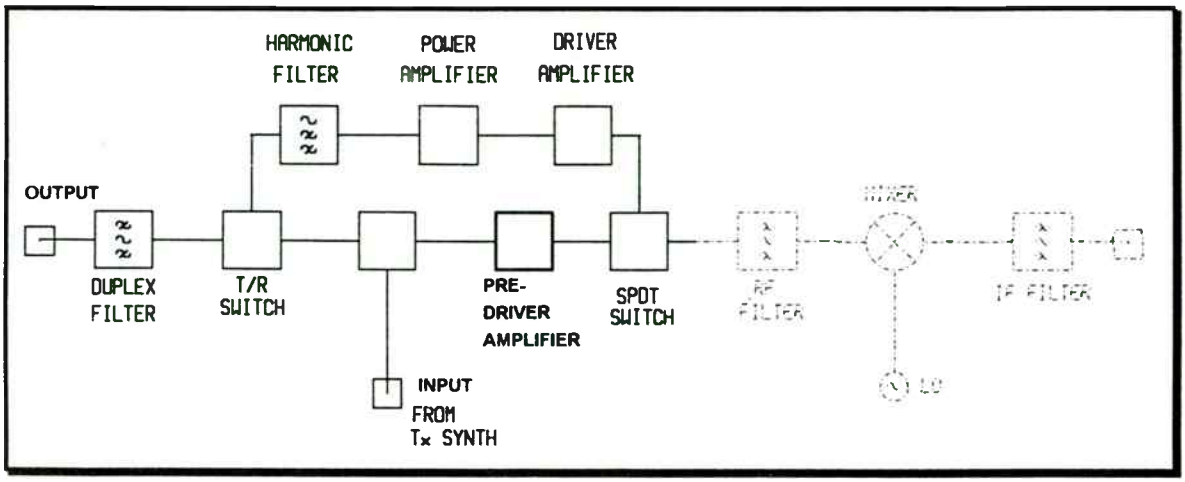
EXAMPLE -- The Low-Noise Amplifier shown aside is used as a small-signal amplifier, on receive, as well as a pre-driver amplifier on transmit. This requires that the amplifier not only have the required GAIN and NOISE FIGURE (simulated using Super-Compact) but also have the necessary 1 dB GAIN COMPRESSION and OUTPUT POWER (simulated using Microwave Harmonica)



S-Parameters of Foundry GaAs MESFET Transistor (Large and Small-Signal Models) and Derived Noise Parameters for the Device at Low-Noise Bias (Raytheon)

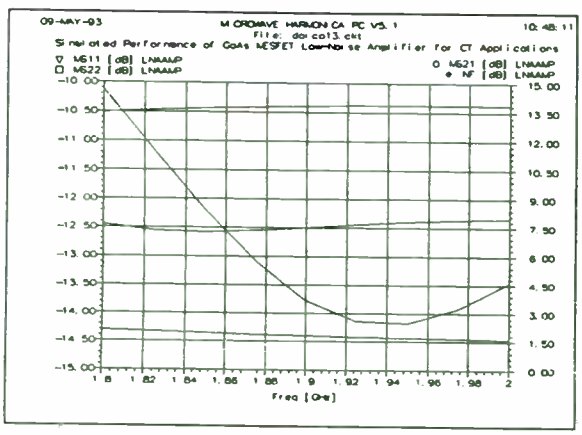


Designing the Low-Noise Amplifier on Receive



Simulating the Low-Noise Amplifier used as a Pre-Driver Amplifier on Transmit

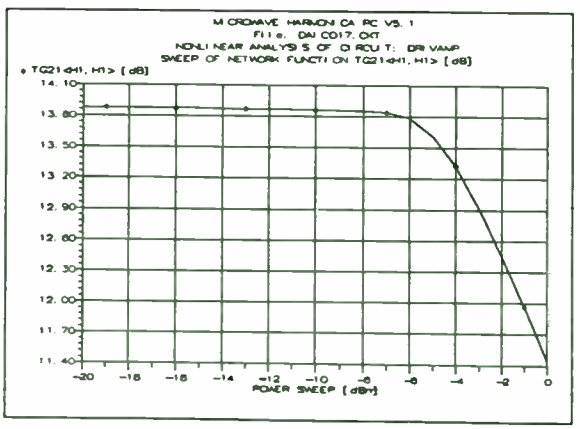
Low-Noise Amplifier

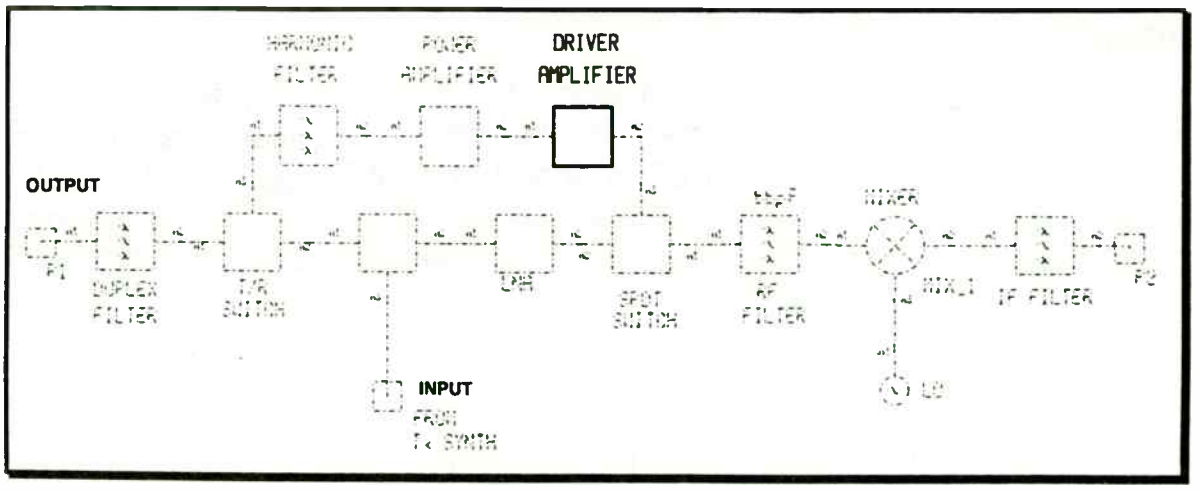


Raytheon MMC-05 400 µm gate width FET used in the low-noise/pre-driver amplifier design. The FET is operated at 25% I_{dss} i.e. at 35 mA. Self-bias used for single-rail operation. Gain is > 13 dB with noise figure of < 2 dB.

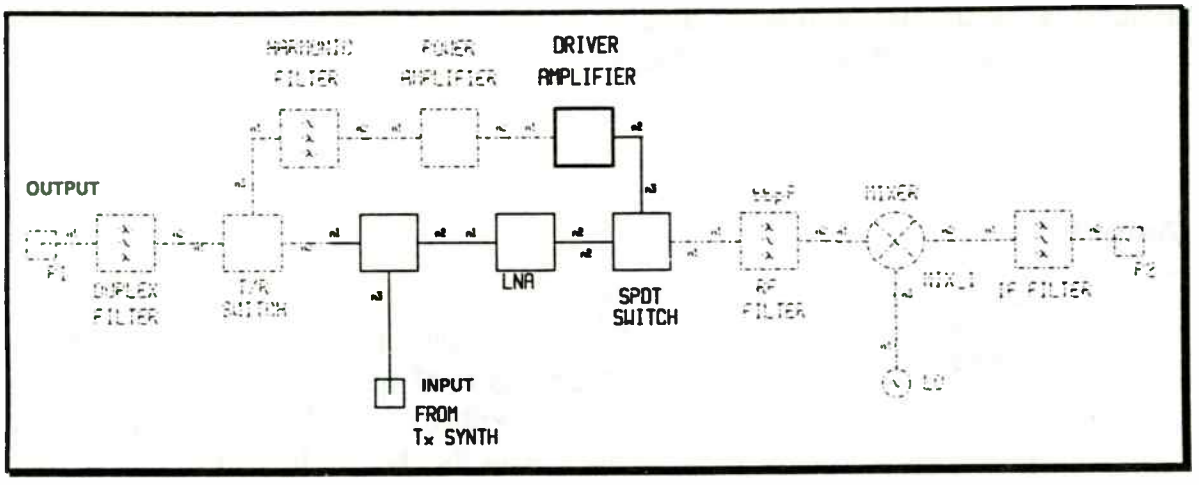
Pre-Driver Amplifier

Used as the pre-driver amplifier on transmit the circuit has a 1 dB gain compression point of + 10 dBm at its output. Power consumption is 70 mW.



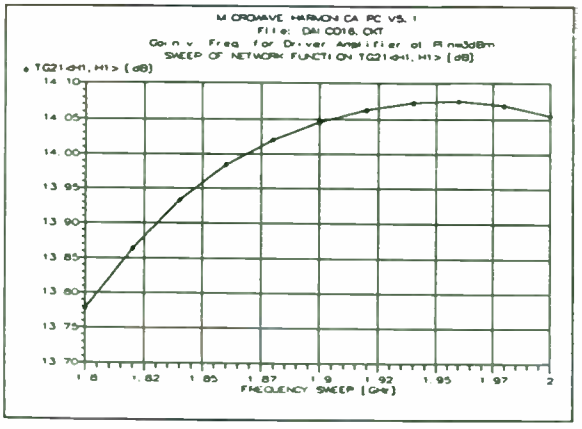


Simulating the Driver Amplifier on Transmit

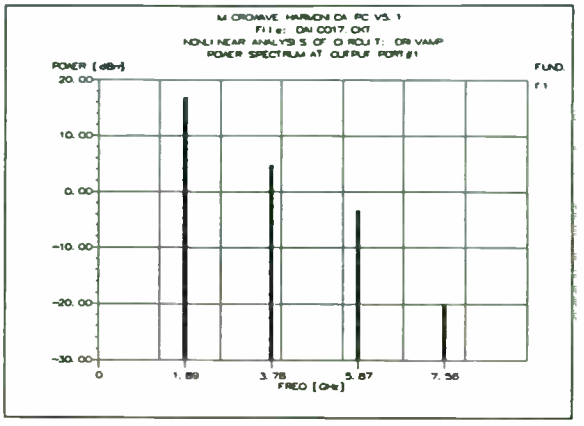


Simulating the Pre-Driver/Driver Amplifier Combination on Transmit

Driver Amplifier

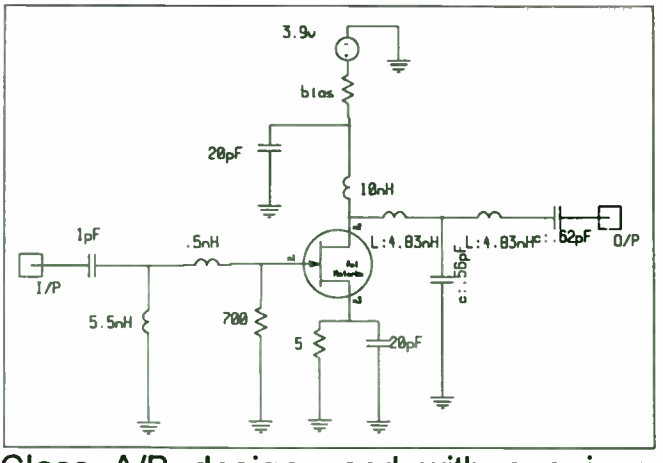


Raytheon MMC-05 400 µm gate width FET used in the driver amplifier design. The FET is operated at 50% I_{dss} i.e. at 70 mA. Self-bias used for single-rail operation. Gain is > 13 dB with noise figure of 2.5 dB.



The harmonic performance of the driver amplifier driven by the output signal from the pre-driver is shown aside. Power consumption is 250 mW. The second harmonic is only 12 dB below the wanted signal -- hence the need for harmonic filtering.

Power Amplifier



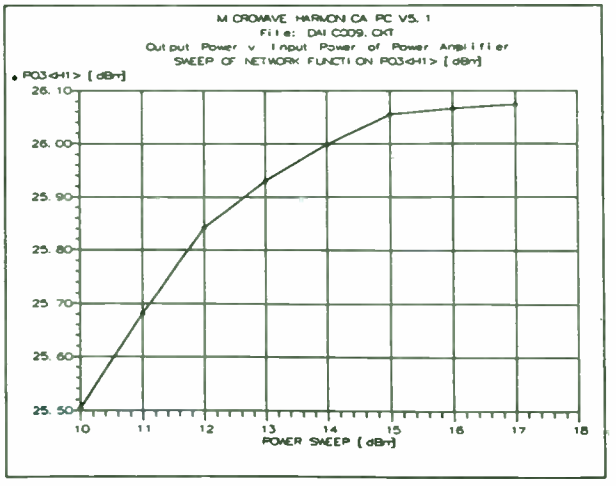
⇦

Power Amplifier required to provide 25 to 26 dBm of output RF power from a single voltage rail that can be as low as 3.3 volts. Need for high DC to RF conversion efficiency.

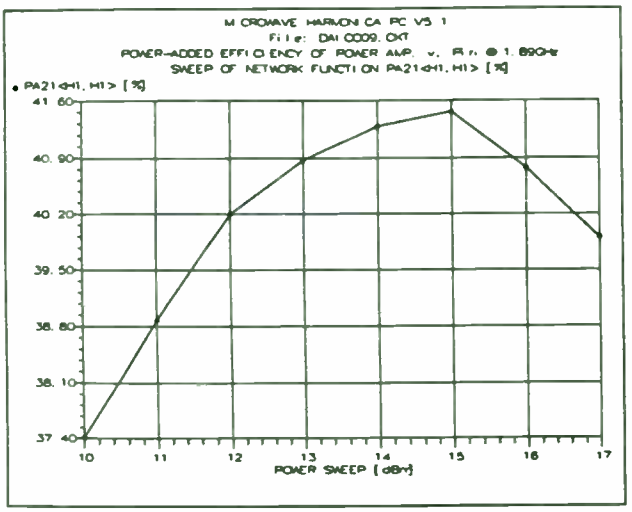
Class A/B design used with a quiescent current of 225 mA.

Power gain of > 12.5 dB achieved at 26 dBm output power.

⇨

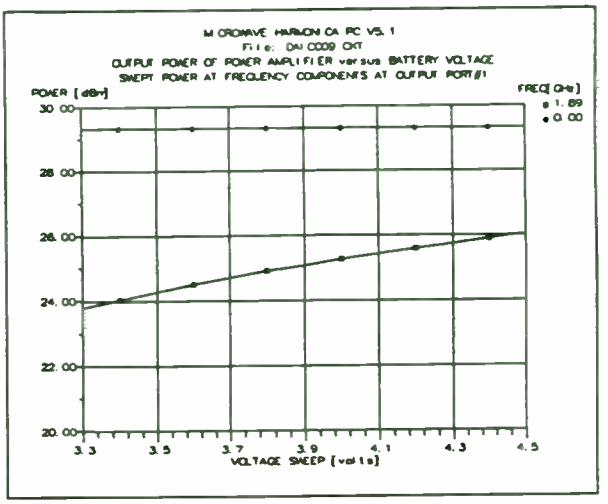


Power Amplifier Design



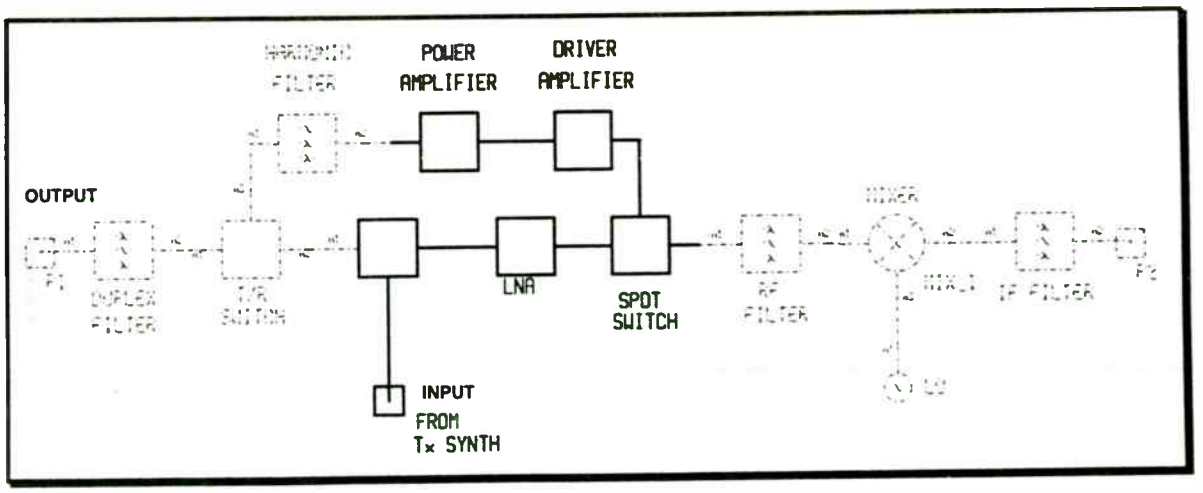
⇦ **Power Added Efficiency**

Power-Added Efficiency is maximized at the required Output Power. DC to RF Conversion Efficiency is close to 40%.



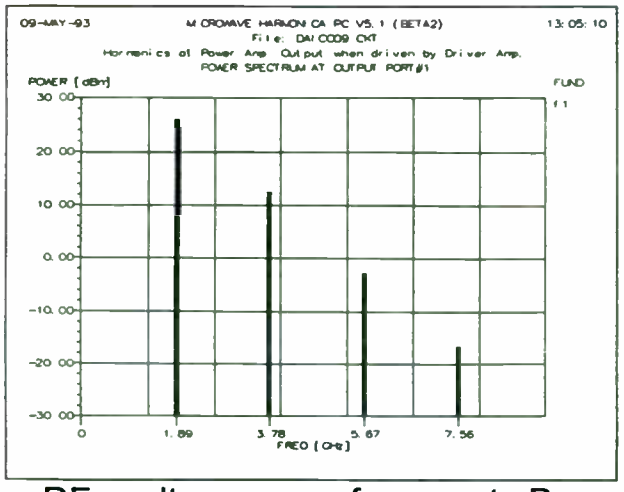
Output Power must vary with battery voltage as little as possible over the "working range" of the battery.

Battery Range ⇨



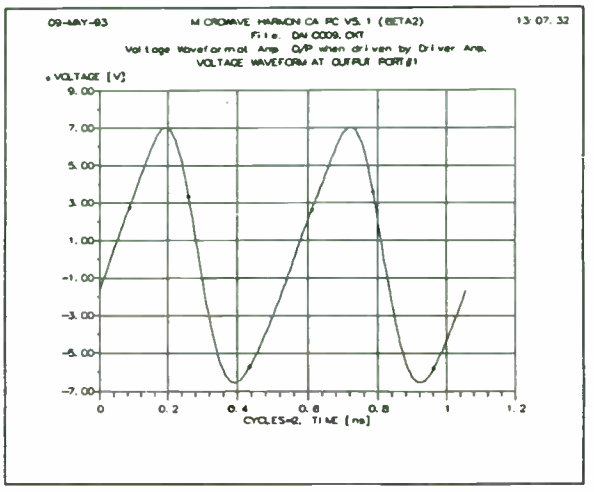
Simulating the Complete CT Transmitter

Power Amplifier Design

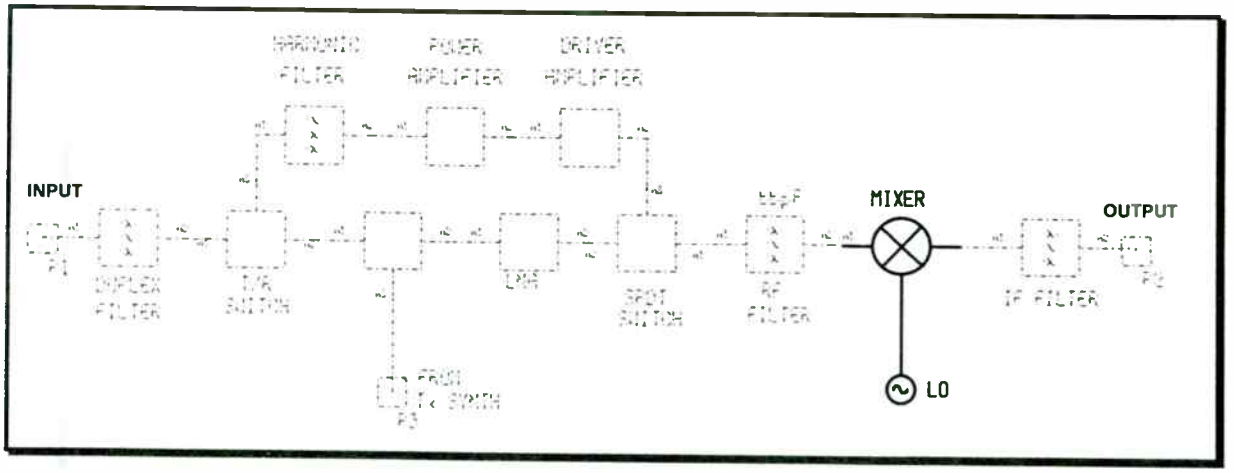


⇦ **Harmonics Generated by Amplifier**

Although Class A/B operation has high DC to RF Efficiency the amplifier generates significant harmonics.

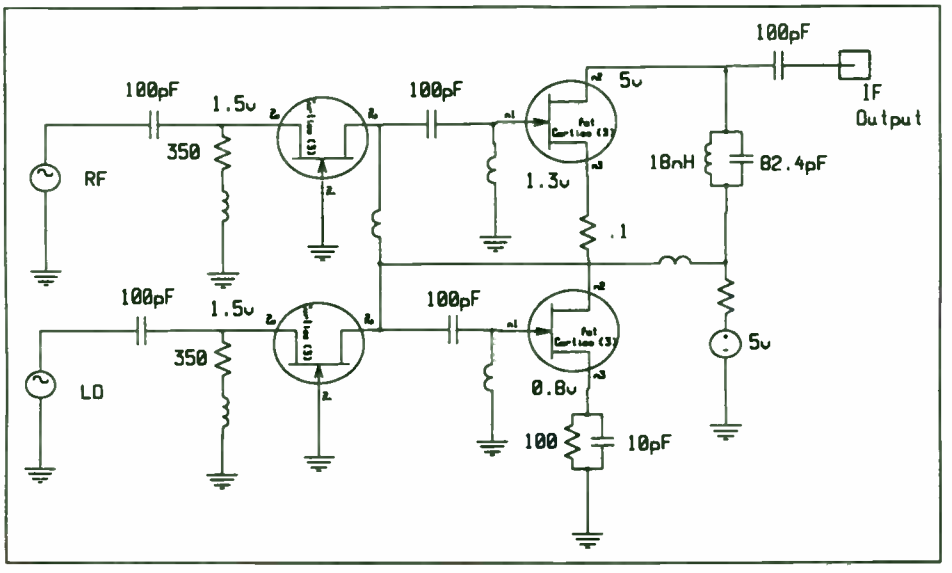


RF voltage waveforms at Power Amplifier Output show high 2nd and 3rd harmonic content. These waveforms are "cleaned up" by the harmonic filter that follows the amplifier. **Spectral Plots** ⇨



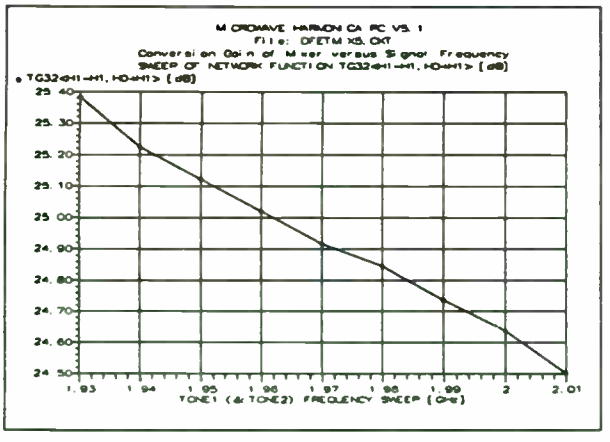
Designing the Mixer on Receive

MESFET Active Mixer Design



- Dual Gate MESFET Mixer
  - Dual-Gate FET is driven by Common-Gate FETs at RF and LO
  - Uses a minimum of passive components to reduce circuit size
  - LO level is -10 dBm -- Average mixer current is 8 mA
  - Mixer has high (25 dB) gain from RF to IF

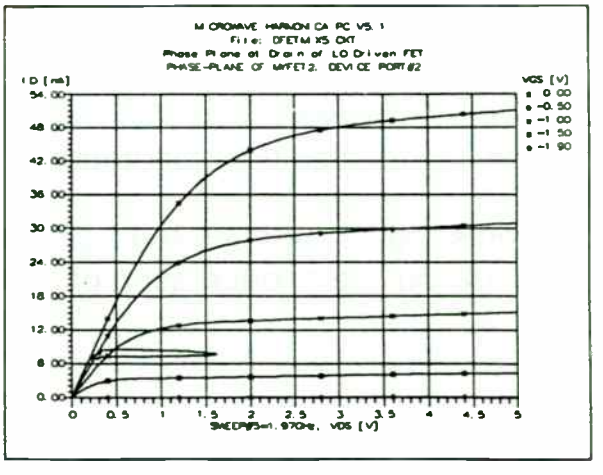
MESFET Active Mixer Design



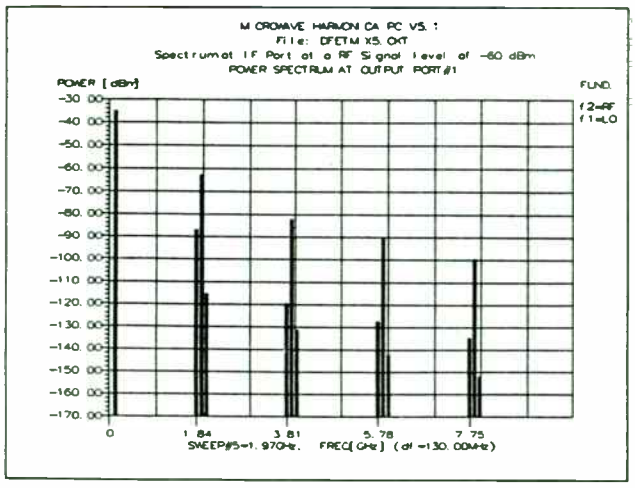
⇦ **RF to IF Conversion Gain v. Signal Frequency**

- Use Microwave Harmonica to design mixer
- Conversion Gain of Mixer varies by only 1 dB over the 1.93 to 2.01 GHz band

Microwave Harmonica allows the investigation of both DC and AC performances. The voltage and current excursions at the drain of the mixer FET are shown aside plotted on top of the I-V charactersitics of the transistor. Current is 8 mA with a voltage swing of 1.5 volts.



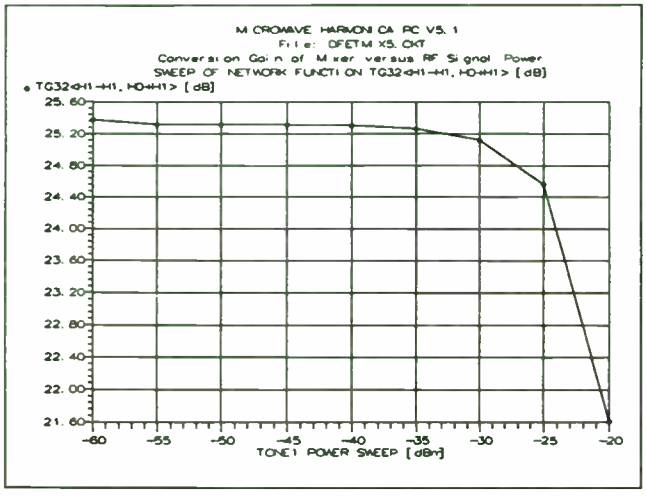
MESFET Active Mixer Design



⇦ **Spectrum at the IF port at an RF signal Level of -60 dBm**

- Rejection of LO and RF signals at the IF port is greater than 50 dB.

Compression of the mixer is plotted aside. The 1 dB compression occurs at -25 dBm RF input power corresponding to a -2.5 dBm (approx. 0.5 mW) output power.

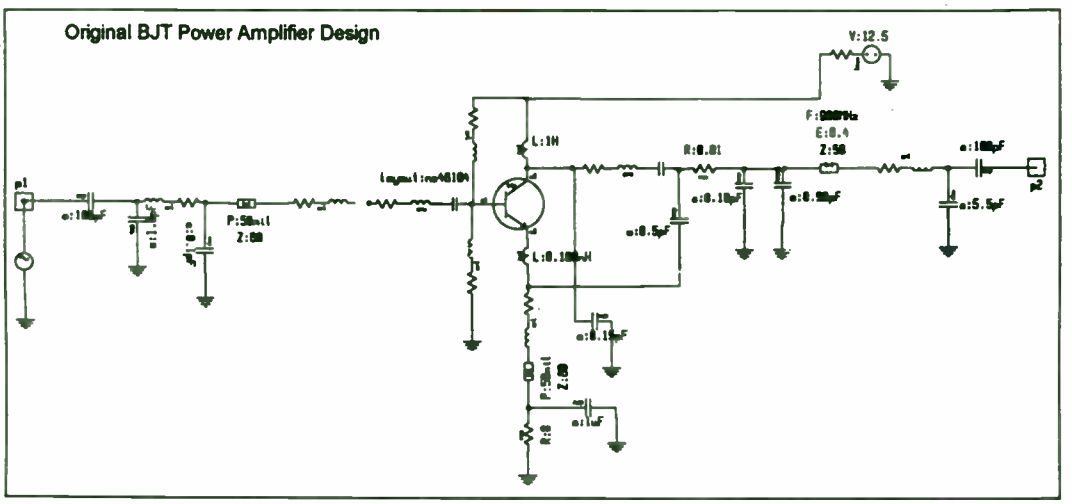


Compression Characteristics of Mixer ⇨

Simulation from Layout is important at RF

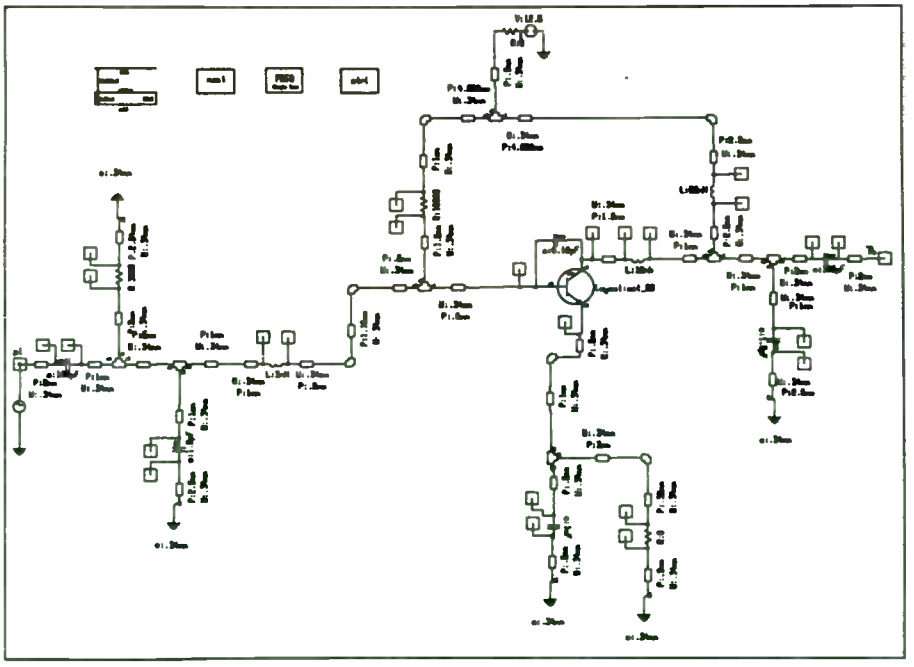
- At low frequencies simulations are usually derived from electrical representations of the circuits
- At higher frequencies the ACTUAL layouts of the circuits influences the performances obtained
  - Layout effects include coupling between components; electrical length of interconnections; parasitic elements; discontinuities
- Schematics used to derive circuit description
  - Provides connectivity check
  - Provides documentation
- Layouts are auto-generated from schematics that contain "layout-linked" information -- e.g. orientation of bends, tees, component footprints, material layers

Simulation from Layout



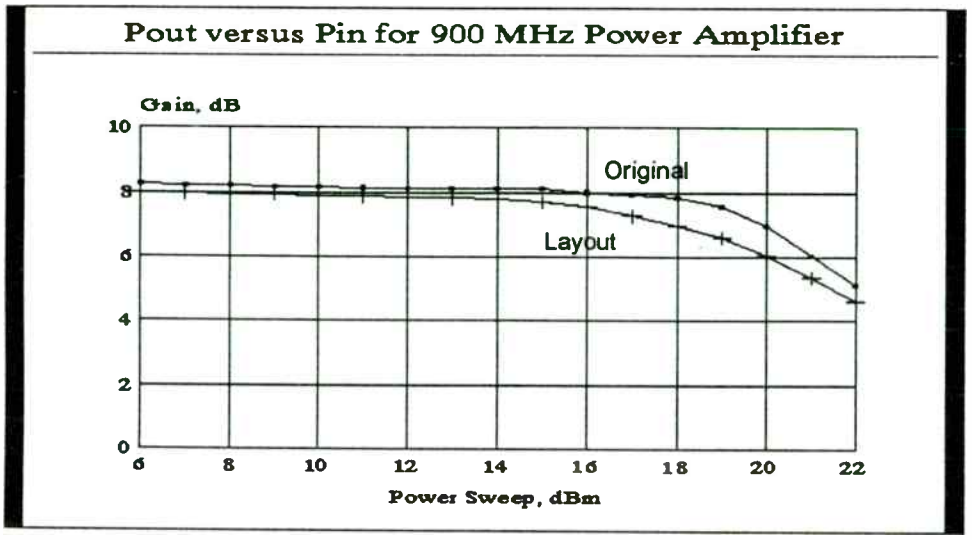
◆ In the original BJT Power Amplifier design no account of layout features are taken. The design consists of simple lumped and distributed elements together with the transistor.

Simulation from Layout



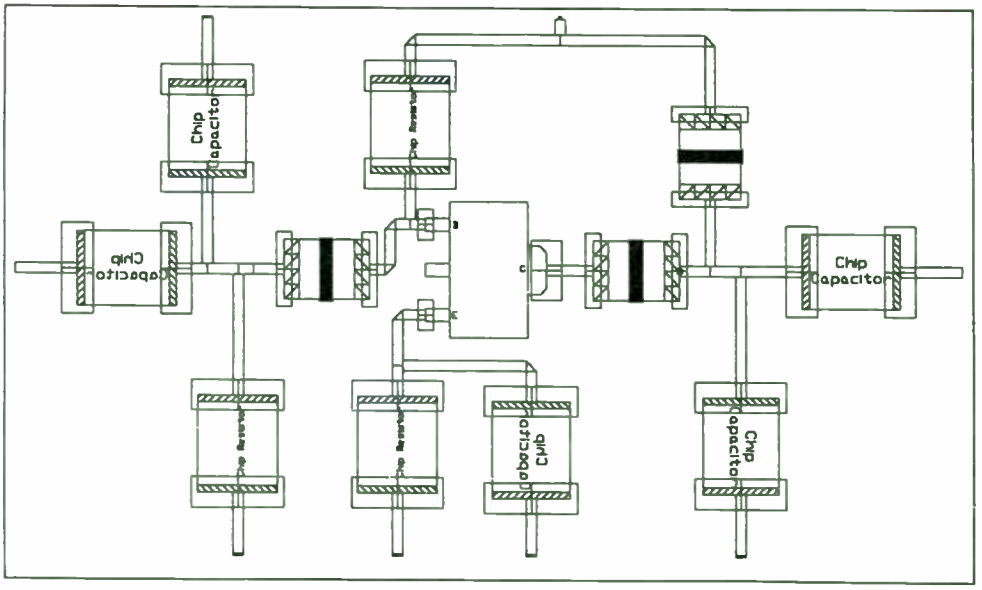
◆ The schematic representing the circuit as laid out on the PCB contains many more elements than the original design including bends, tees, bond pads etc..

Simulation from Layout



- The above response shows the effect circuit layout has on the gain compression charactersitics of a Class A BJT Power Amplifier at 900 MHz

Simulation from Layout



- The BJT Amplifier Layout is derived automatically from the circuit schematic. The amplifier shown above consists entirely of surface mounted devices. Individual layers for solder attach, drill holes, transmission lines are created.

Additional CAD Tools for RF Design

- **Time Domain Simulation with RF/Microwave Components**
  - *Conventional SPICE programs only contain "low frequency" models*
  - *To overcome these problems new versions of SPICE are being introduced that contain microwave component models:*
    - *Models developed using the Method of Lines*
    - *Models that are "translated" from the frequency to the time domain using "convolution"*
  - *Compact Software has produced new time domian simulator called Super-Spice® using an X-Windows/Motif Interface on Workstations*
  - *Multi-layer circuits such as multi-level PCBs and MCMs can be analyzed. Transient as well as steady-state analysis can be performed*
  - *Super-Spice allows RF design engineers to investigate mixed-signal sub-systems e.g. Phase-Locked Loops at high RF frequencies*

Additional CAD Tools for RF Design

- **Electromagnetic Simulation**
  - *Many circuit structures cannot be analyzed using conventional circuit simulators*
  - *To overcome these problems ElectroMagnetic (EM) Simulators based on the Method of Moments have been developed*
  - *Compact Software has produced an efficient EM simulator called Explorer® using an X-Windows/Motif Interface on Workstations*
  - *Multi-layer circuits such as multi-level PCBs and MCMs can be analyzed. S-Parameter data is then transferred to Super-Compact and/or Microwave Harmonica for incorporation in other sub-systems*
  - *Today, EM simulators are much slower than nodal-based circuit simulators – that situation will change with new mathematical techniques and faster CPUs*

◆ Conclusions

- CAD Tools for the Design of Circuits and Systems at RF are Commercially Available and Mature Products
- There is a Continuing Growth in the Supply of Component Libraries for both Circuit and System Level Simulation
- Products for the Layout of Circuits and Systems at RF with Direct Links to Performance, Yield and Manufacturing Costs are being Developed and some are Available Commercially
- Examples have been given of Circuit and System Level Designs and the Importance of New CAD Tools such as Electromagnetic Simulation